The Biometric Society

BIOMETRICS

FOUNDED BY THE BIOMETRICS SECTION OF THE AMERICAN STATISTICAL ASSOCIATION

TABLE OF CONTENTS

Number 3 September 1954 Volume 10

Material for *Biometrics* should be addressed to Miss Gertrude Cox, Institute of Statistics, Box 5457, Raleigh, North Carolina, except that authors residing in one of the following organized regions can expedite the handling of their papers by submitting them to the Assistant Editor for that region.

British Region: Dr. D. J. Finney, Dept. of Stat., Univ. of Aberdeen, Aberdeen, Scotland; **Australasian Region**: Dr. E. A. Cornish, University of Adelaide, Adelaide, Australia; **French Region**: Dr. Georges Teissier, Faculte des Sciences de Paris, 1 rue V. Cousin, Paris, France.

Material for Queries should go to Professor G. W. Snedecor, Statistical Laboratory, Iowa State College, Ames, Iowa.

Articles to be considered for publication should be submitted in triplicate.

The Biometric Society is an international society devoted to the mathematical and statistical aspects of biology and welcomes to membership biologists, mathematicians, statisticians and others who are interested in its objectives. Through its regional organizations the Society sponsors regional and local meetings. National secretaries serve the interest of members in Denmark, the Netherlands, India, Germany, Japan, Sweden and Brazil and there are many members "at large". Dues in the Society for 1954 for residents of the Western Hemisphere are as follows: Full membership including subscription to *Biometrics* is $7.00. Members of the Biometrics Section of the American Statistical Association who subscribe to the journal through that organization may become members of The Biometric Society on the payment of $3.00 annual dues. For members in other parts of the world, full membership including subscription to *Biometrics* is $4.50, except that members who subscribe to the journal through the American Statistical Association pay annual dues of $1.75. Information concerning the Society can be obtained from the Secretary, The Biometric Society, Drawer 1106, New Haven 4, Connecticut, U.S.A.

Annual subscription rates to non-members are as follows: For American Statistical Association Members, $4.00; for subscribers, non-members of either American Statistical Association or The Biometric Society, $7.00. Subscriptions should be sent to the Managing Editor, *Biometrics*, P. O. Box 5457, Raleigh, North Carolina, U.S.A.

Entered as second-class matter at the Post Office at New Haven, Conn., under the Act of March 3, 1879. Additional entry at Richmond, Va. Business Office, 52 Hillhouse Ave., New Haven, Conn. *Biometrics* is published quarterly—in March, June, September and December.

DECISION BETWEEN TWO ALTERNATIVES—HOW MANY EXPERIMENTS?*

P. M. Grundy, D. H. Rees and M. J. R. Healy
Rothamsted Experimental Station

When in the course of scientific research a new technological process or agricultural treatment is suggested, the question at once arises whether the change-over from established practice is going to be economically justified. The answer to this question can be found by comparing the extra cost of the new process with the value of the extra output due to its adoption, and this latter quantity will have to be estimated by a set of experiments. Now experiments themselves cost money, and we are faced with a compromise; we can make certain of getting the right answer to our question at the expense of a very large experimental programme, or we can economise on our experiments at the risk of arriving at a wrong conclusion. It seems clear that there will be an optimum amount of experimentation which will in some sense give us the best return for our money, and this paper is concerned with the determination of such an optimum.

The extra returns due to introducing the new process will depend on the scale on which it will be applied and on the increase in output per unit over the original process, and knowing these we could estimate the expected loss due to wrong decisions and balance it against the cost of experimentation. The increase in output per unit, however, is only determined from the results of the experiments themselves, and there are obvious advantages in some kind of sequential scheme in which the decision on whether or not to continue experimenting depends on the results so far obtained. With the present approach, the investigation of a fully sequential technique presents rather formidable difficulties, although some progress has been made in this direction; but a fairly detailed solution to the problem is possible under conditions analogous to *double sampling* as used in quality control. We suppose that a preliminary experiment is carried out, and that in the light of this we decide to adopt or reject the new process or else to postpone our decision

*Read at the Third International Biometric Congress, Bellagio, September, 1953.

until we have done one further set of experiments whose size depends on the results so far obtained. This is likely to be less efficient than a fully sequential scheme, but is of practical importance when a decision is needed quickly, so that there is only time to carry out two batches of experiments. A full mathematical treatment of the problem will be given in a forthcoming paper, the results of which are summarised here.

We suppose, then, that the new process increases the output per unit by a quantity η and that the net gain due to introducing the new process* is given by

$$\mu = k'(\eta - c)$$

where k' is a known positive factor depending on the scale of application and c is a known constant allowing for the difference in cost between the old and new processes and the capital cost of the change. The change-over will then be economically justifiable if $(\eta - c)$ is positive. To estimate η we carry out a set of experiments which provide a set of independent estimates y, normally distributed with standard deviation σ. (We assume that σ is known with sufficient accuracy from previous experience, and that it includes all sources of variation that affect the y's.) Then, if we carry out n experiments, our decision as to the adoption or rejection of the new process will be based on the sign of $(\bar{y} - c)$, where $\bar{y}$ is the mean of the n estimates y, and we can calculate the probability P of accepting the new process as a function of η. If the cost of each experiment is k, then the total cost of our experiments is kn and the expectation of gain (measured from the status quo) allowing for the uncertainty of our information is $P\mu$, so that the total of costs plus losses is

$$R = kn - P\mu$$

which may be referred to as the *total risk*.The optimum value of n will be that which minimises R. .However, we have still to face the difficulty that P, and hence R, is a function of the unknown η.

In the present approach, we carry out a preliminary experiment and obtain an estimate y_1 . P is taken to be the conditional probability of accepting the new process, given y_1 . If we now average R over the fiducial distribution of η, which is normal with mean y_1 and standard deviation σ, we arrive at an *averaged risk* $\bar{R}$.** We choose the value of

*If the gain is to continue over a period of years, a solution can be obtained by using an equivalent capital sum.

**This approach seems intuitively reasonable, and results given below indicate that it provides a satisfactory solution to the present problem; but we do not wish to claim here any general optimal properties for it.

n to minimise this quantity. On the present assumptions, this averaged risk takes the form

$$\bar{R} = kn - k'(y_1 - c)\phi\left[\frac{y_1 - c}{\sigma}\sqrt{\frac{n}{n-1}}\right] - k'\sigma\sqrt{\frac{n-1}{n}}\,\phi'\left[\frac{y_1 - c}{\sigma}\sqrt{\frac{n}{n-1}}\right]$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{(-1/2)t^2}\,dt.$$

The possible modes of behaviour of $\bar{R}$ as a function of n are illustrated diagramatically in Fig. 1. As n increases from 1, $\bar{R}$ may increase

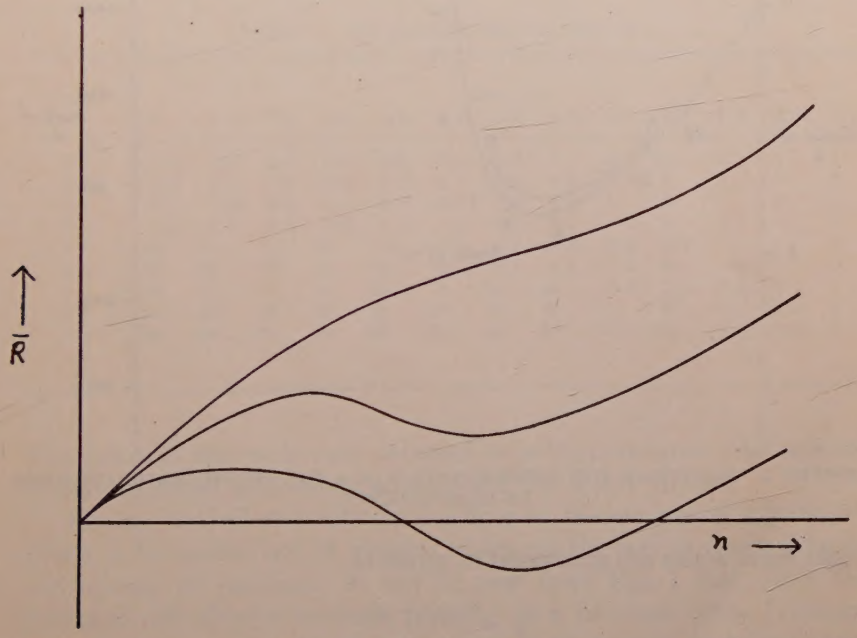


FIGURE 1. AVERAGED RISK AS A FUNCTION OF THE NUMBER OF EXPERIMENTS

steadily or else may pass through a maximum and then a minimum value, thereafter increasing steadily. Furthermore, the minimum value may be greater than the value at $n = 1$. It is clear that only when the minimum value is *less* than the value at $n = 1$ is it worth while doing more experiments (the optimum number of experiments being given by the value of n at the minimum); otherwise, the decision to

adopt or reject the new process should be based on the result of the first experiment alone.

The value of n which minimises $\bar{R}$ is actually given by the larger root of the equation

$$n^{3/2}(n-1)^{1/2} = \frac{1}{2}\frac{k'\sigma}{k}\phi'\left[\frac{y_1 - c}{\sigma}\sqrt{\frac{n}{n-1}}\right]$$

This can be solved iteratively for given values of the constants; but it is much simpler to use the nomogram shown in Fig. 2. The left and

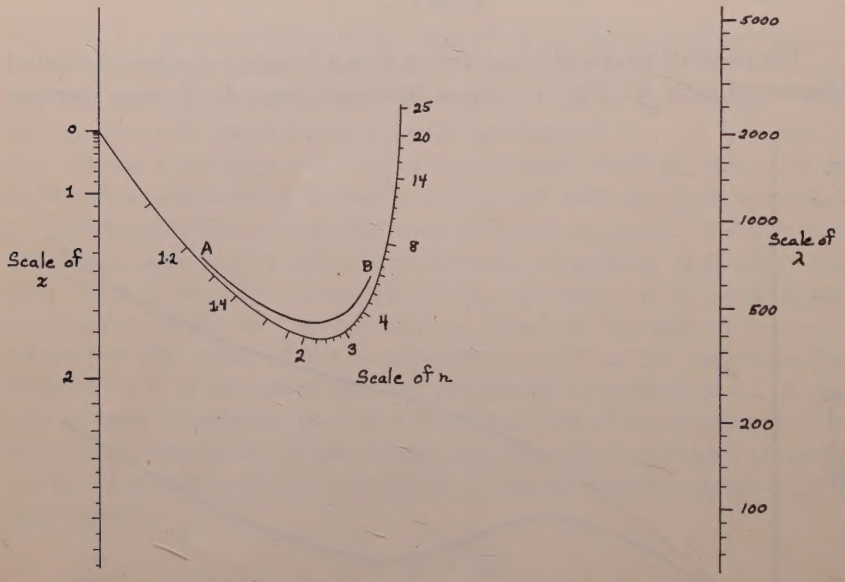


FIGURE 2. NOMOGRAM FOR DETERMINATION OF n, THE RECOMMENDED NUMBER OF EXPERIMENTS

right hand scales are graduated in terms of

$$x = \left|\frac{y_1 - c}{\sigma}\right|$$

and

$$\lambda = \frac{k'\sigma}{k},$$

and a line joining the appropriate points on these scales intersects the curved scale at points corresponding to the two roots of the above equation. We have still to ensure that the minimum value of $\bar{R}$ is actually less than the value at $n = 1$. The condition for this to be true

can be put in the simple form

$$\frac{\xi\phi(-\xi)}{\phi'(\xi)} < 1 - \frac{1}{2n},$$

where

$$\xi = \sqrt{\frac{n}{n-1}} \left| \frac{y_1 - c}{\sigma} \right|$$

and this has been incorporated in the nomogram in the form of the ungraduated curve AB. If the line joining the outer scales fails to intersect or to pass above this curve, the minimum $\bar{R}$ is greater than the value at $n = 1$ and no further experimentation is recommended.

The recommended n can, alternatively, be obtained from Table 1.

TABLE 1. RECOMMENDED VALUES OF n

log λ	Value of x												
	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
2.0	4.8	4.7	4.5	4.3	3.9	3.4	2.9	1	1	1	1	1	1
2.2	5.9	5.8	5.7	5.3	4.9	4.3	3.7	3.0	1	1	1	1	1
2.4	7.3	7.3	7.0	6.7	6.2	5.5	4.8	4.0	1	1	1	1	1
2.6	9.2	9.1	8.8	8.3	7.7	7.0	6.1	5.2	4.1	1	1	1	1
2.8	12	11	11	10	9.7	8.8	7.7	6.6	5.4	4.1	1	1	1
3.0	14	14	14	13	12	11	9.7	8.4	7.0	5.5	4.0	1	1
3.2	18	18	17	16	15	14	12	11	8.9	7.2	5.5	1	1
3.4	23	23	22	21	19	17	15	13	11	9.3	7.3	5.3	1
3.6	28	28	27	26	24	22	20	17	14	12	9.5	7.2	1

The values in this table were obtained from the nomogram, this method being considered sufficiently accurate for the object in view.

Having arrived at a rule for deciding on the amount of experimentation to be carried out, it remains to discuss the way in which the rule will operate in practice. It can be seen from Fig. 1 that, since the averaged risk always increases initially as n increases from 1 (except when $x = 0$ exactly), there is always a small amount of extra experimentation which is definitely uneconomic. This is very reasonable, since the only reason for doing more experiments is the possibility that they may contradict the initial result, and a fair amount of data will be necessary before we shall be prepared to go against our original findings. Fig. 3 shows the probability distribution of n, the optimum number of experiments, for $\lambda = 10^{2.5}$ and $\theta/\sigma = 0$, 1.5 and 3, where $\theta = \eta - c$. Notice that whatever the result of the first experiment there is an upper

limit (depending on λ) to the number of experiments that can be recommended.

What the rule accomplishes can best be appreciated in terms of the

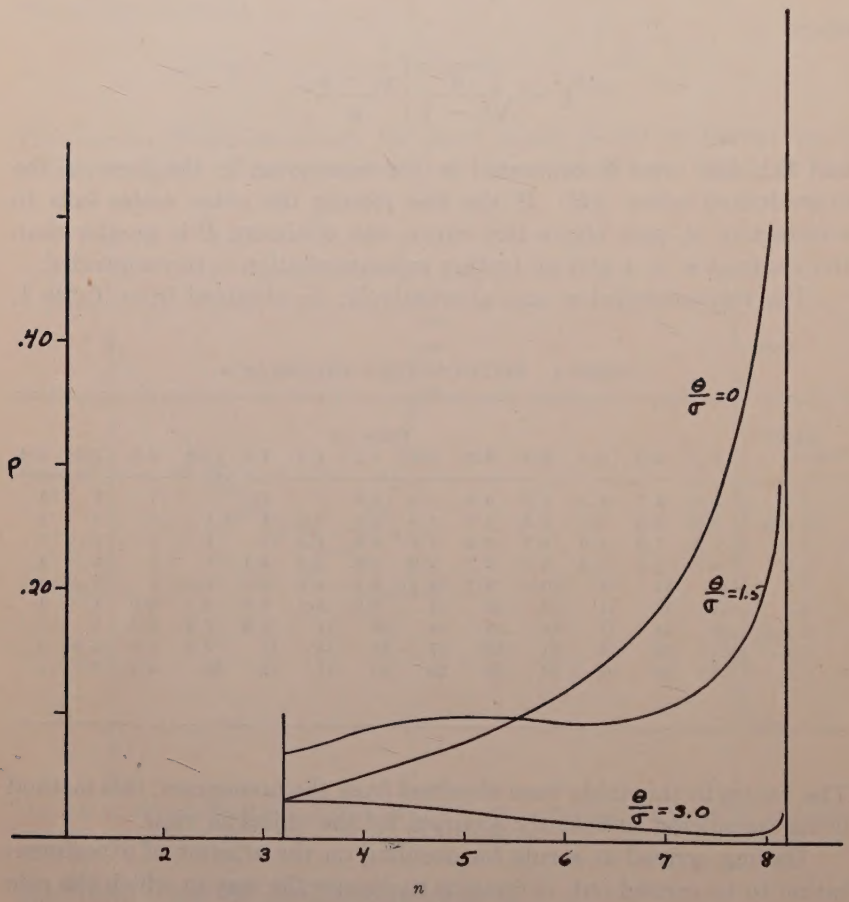


FIGURE 3. FREQUENCY DISTRIBUTION OF $n > 1$ FOR $\lambda = 10^{2.5}$

expected cost plus losses under different circumstances. In Fig. 4, this total expected loss is plotted against θ/σ for λ equal to $10^{2.5}$ and 10^3. Also shown are the total losses for two "single sampling" schemes, where the amount of experimentation is decided once for all. The amounts considered are

(i) $n = 1$;

(ii) $n = \frac{3}{4}n_{\max}$,

where n_{max} is the greatest number of experiments ever recommended under the double sampling rule. The second scheme gives $n = 6.2$ for $\lambda = 10^{2.5}$ and $n = 10.8$ for $\lambda = 10^3$. The first scheme is uneconomical unless the effect of the new process is very small, when our potential gain to set against the cost of experimentation is small, or very large,

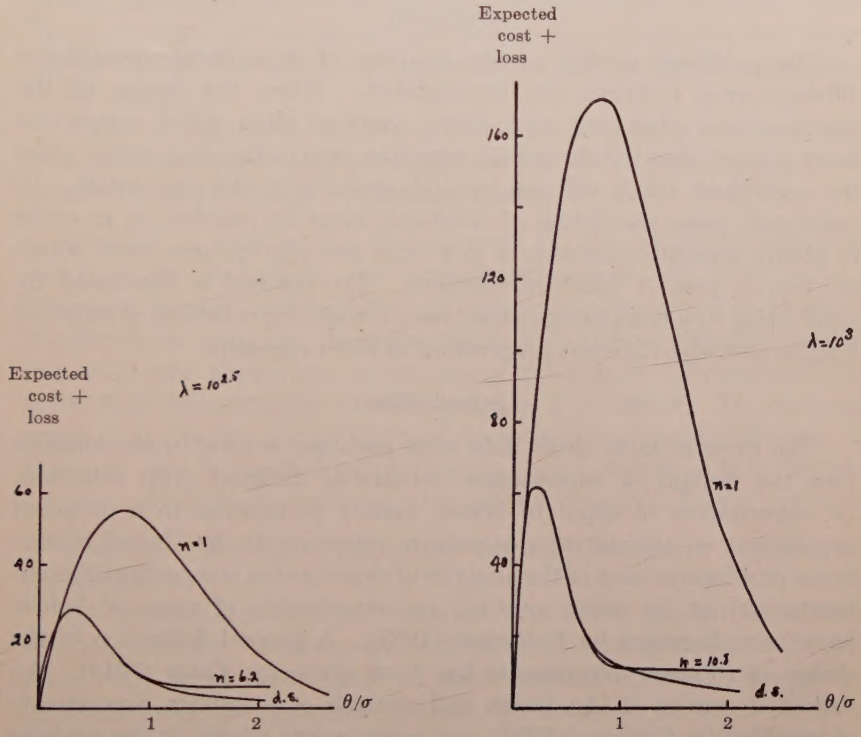


FIGURE 4. EXPECTED (COSTS + LOSSES) FOR TWO SINGLE SAMPLING SCHEMES, $n = 1$ AND $n = \frac{1}{4}n_{max}$ AND FOR DOUBLE SAMPLING

when a correct decision will nearly always be made; the second tends to be wasteful, especially when the effect of the new process turns out to be large. It is worth mentioning that no scheme on a single or double sampling basis can give better results than the present rule for *all* values of θ/σ.

The original idea behind this research was given by Dr. F. Yates in an article in *Nature* (*170*, 138, 1952) where he gave a solution to the quantitative problem in which the optimum level of some treatment is to be determined.

THE ANALYSIS OF EXPERIMENTS CONTAINING DIFFERENT CROP ROTATIONS

F. YATES
Rothamsted Experimental Station.

Summary

The problems arising in the analysis of experiments containing different crop rotations are investigated. When the design of the experiment is such that each block contains plots which sometimes carry a given crop but do not all carry the crop in the same set of years the year-block totals will not be orthogonal with the plot totals. In most such cases the fitting of constants must be resorted to in order to obtain separate estimates of plot error and plot × year error which are free of year × block interactions. The method is illustrated by application to a rice-pasture experiment containing rotations of different lengths and with different proportions of rice to pasture.

Introduction

The present paper deals only with problems arising in the analysis (not the design) of experiments containing different crop rotations, i.e. experiments of type (b) below, mainly in relation to a proposed experiment on alternative rice-pasture rotations in the United States. Some problems arising in the analysis of experiments comparing different treatments on the same rotation, i.e. experiments of type (a) below, have been discussed by Patterson (1953). A general discussion of the design of rotation experiments has been given by Yates (1949). An earlier discussion of the design and analysis of long-term experiments is provided by Cochran (1939), and some points arising in the analysis of experiments of type (b) are considered by Crowther and Cochran (1942).

Terminology

The design and analysis of experiments involving crop rotations is one that has received relatively little notice in the literature, and a brief note on the terminology of the subject may therefore be helpful to the understanding of this paper.

Sequence. In this paper the term sequence is used to denote a sequence of crops, treatments or crop-treatment combinations which differs in any respect from other such sequences occurring in the experiment.

Cycle, period, phase. These terms are used in their customary sense. If we have a repetitive sequence of crops or treatments c_1 , c_2 , $\cdots$, c_n , c_1 , c_2 , $\cdots$, c_n a single repetition is termed a cycle. The number of years (or other time units) in the cycle is termed its period. Sequences starting at different points in the same cycle are said to be in different phase. There are thus n possible phases of a cycle of period n. The term phase is also used to denote the different components of the cycle. These alternative meanings are not likely to cause confusion and follow common usage (e.g. phases of an alternating electric current and phases of the moon).

If two or more cycles of differing period are included in the experiment then the whole experiment will follow a cycle whose period is the lowest common multiple of the two or more periods.

Rotation. A rotation is a definite cycle of crops grown in successive years on the same land. In agricultural practice the term is used somewhat loosely, and includes not only minor variations in cropping from cycle to cycle, e.g. the substitution of one cereal crop for another, but often also alterations in the length of the cycle. In experimental work such variations are naturally kept to a minimum. The separate crops of a rotation are in agricultural terminology called courses. If the same crop occurs twice in a rotation it will constitute two different courses. An n course rotation is therefore a cropping cycle of period n.

Rotation experiments. There are two main classes of rotation experiments:

(a) Experiments on the effects of treatments applied to a fixed rotation of crops. The experimental treatments may be repeated year after year, or may be varied in some manner which is regarded as appropriate to the questions at issue. A common device is to use a cycle of treatments. If the cycle of treatments on a given plot of a rotation experiment has the same period as the cropping cycle a given crop will have the same treatment each time it is grown. If the periods are different the treatment will vary in a cyclic manner, with the important consequence that any given combination of crop and treatment-phase occurs on different plots in successive crop cycles, thereby considerably increasing the accuracy.

(b) Experiments comparing the effects of different rotations. Here the different crops themselves act as treatments, and plots between which comparisons have to be made will not always be carrying the same crop in the same year. This introduces the additional problem of design of ensuring that the plots to be compared simultaneously carry the same crop in a sufficient number of years for the necessary com-

parisons to be made, a problem which is particularly troublesome when the effect of rotations of different length is under investigation.

Experiments of class (b) may, and often do, contain other experimental treatments in addition to the variations in cropping.

Series. In rotation experiments of type (a) the area of land is usually divided into separate parts for the different crop phases. Except when the same crop occurs more than once in the cycle each part will in any one year carry a different crop. These separate parts are termed series.

Blocks. The whole experiment, or if in series the separate series, will normally be divided into blocks, as in ordinary one-year agricultural experiments. Each block may contain a complete replicate of all the sequences in the experiment or the series, or the device of confounding may be used to reduce block size, in which case each block will only contain part of a replicate.

Preliminary years. One, two or more years at the beginning of a cyclical experiment usually have to be excluded from the main analysis because treatments which would have been applied had the experiment been started earlier will not in fact have been applied. These years are known as the preliminary years. Their number will depend on the nature of the treatments and other factors.

Phase differences. In rotation experiments many different types of contrast may be of interest. One type, which has no analogy in one-year experiments, is the contrast between different phases of the same cycle. Such contrasts may be termed phase differences. In a crop rotation containing the same crop twice in the rotation, for instance, there will be a phase difference for this crop. This phase difference can only be estimated with precision if the phases carrying the same crop occur in the same blocks. Thus with the crop rotation PQPR estimation of the phase difference requires arrangement in two series instead of four. With the rotation PPQR a single series must be used if estimates are to be available each year.

The problem

At the Summer Statistics Conference organized by the North Carolina Institute of Statistics in 1952 and held at Blue Ridge, North Carolina, an experiment to compare various rotations of rice and grass pasture was discussed. Comparisons were required between the rotations:

A. 1 year rice, 2 years grass
B. 2 years rice, 2 years grass
C. 1 year rice, 3 years grass

The design proposed was to compare all phases of these three rotations in three randomised blocks of 11 plots each. A complete cycle of this experiment requires 12 years, after 3 preliminary years. One replicate of such a cycle (excluding the preliminary years) is shown in Table 1.

TABLE 1. ONE REPLICATE OF THE RICE-PASTURE ROTATION EXPERIMENT

Year	A letter denotes the plot is carrying rice. A			B				C				Total
	1	2	3	4	5	6	7	8	9	10	11	
1	a_1			b_1			b'_1	c_1				Y_1
2		a_2		b'_2	b_2				c_2			Y_2
3			a_3		b'_3	b_3				c_3		Y_3
4	a_4					b'_4	b_4				c_4	Y_4
5		a_5		b_5			b'_5	c_5				Y_5
6			a_6	b'_6	b_6				c_6			Y_6
7	a_7				b'_7	b_7				c_7		Y_7
8		a_8				b'_8	b_8				c_8	Y_8
9			a_9	b_9			b'_9	c_9				Y_9
10	a_{10}			b'_{10}	b_{10}				c_{10}			Y_{10}
11		a_{11}			b'_{11}	b_{11}				c_{11}		Y_{11}
12			a_{12}			b'_{12}	b_{12}				c_{12}	Y_{12}
Total	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	R

The yields of rice in the three rotations can be assessed by comparing the means of the a, $\frac{1}{2}(b + b')$ and c yields, and the difference between the first and second year's rice in the B rotation (which we may term the crop phase difference) can be assessed by comparing the means of the b and b' yields. All these comparisons will be free of year differences since each year is equally represented in each mean. The first group of comparisons will involve plot differences, whereas the phase comparisons will only involve plot and year interaction. Similar comparisons can be made over a shorter period than 12 years, though some of the symmetry is then lost. If only six years results are available, for example, each of the A plots is represented twice, but of the B plots plot 4 is represented four times, plots 5 and 7 three times, and plot 6 twice, and similarly with the C plots.

The problem is to estimate the errors of these and other comparisons that may require to be made. The ordinary subdivision of the analysis

of variance into a part derived from the totals of plots over all years, and another part (further sub-divided if necessary) derived from the plots × years interaction, breaks down, since the different plot totals involve year differences.

The full analysis of an experiment of this type provides an interesting example of the partition of degrees of freedom in material which possesses a certain degree of balance and orderliness but is not fully orthogonal. In order to elucidate the various points at issue we will first consider the analysis of two somewhat simpler types of experiment.

A rotation with all crops different, treatments tied to crops

As a simple example we may take a three course rotation experiment to compare two treatment cycles (or fixed treatments), with three replicates of each of the three phases. Denote the treatment cycles by A and B. It is assumed that a given crop always has the same treatment i.e. that the periods of the crop and treatment cycles are the same. Thus we might have an experiment in which treatment cycle A was the application of farmyard manure to the potato crop in a potatoes-barley-wheat rotation, treatment cycle B being the control (no farmyard manure). The barley and wheat crops will then measure the first and second year residual effects of the farmyard manure.

As all the crops are different each phase of the rotation will normally be grouped in a separate series of three blocks. Table 2 shows the 6

TABLE 2. CROP TREATMENT SEQUENCES IN A THREE COURSE ROTATION WITH ALL CROPS DIFFERENT

(Crops: P, Q, R; treatment cycles: A_1, A_2, A_3; B_1, B_2, B_3)

	Series:	I		II		III	
	Blocks:	1, 2, 3		4, 5, 6		7, 8, 9	
	Sequence:	1	2	3	4	5	6
Preliminary years	−1	$Q\,A_2$	$Q\,B_2$	$P\,A_1$	$P\,B_1$	$R\,A_3$	$R\,B_3$
	0	$R\,A_3$	$R\,B_3$	$Q\,A_2$	$Q\,B_2$	$P\,A_1$	$P\,B_1$
Experimental years	1	$P\,A_1$	$P\,B_1$	$R\,A_3$	$R\,B_3$	$Q\,A_2$	$Q\,B_2$
	2	$Q\,A_2$	$Q\,B_2$	$P\,A_1$	$P\,B_1$	$R\,A_3$	$R\,B_3$
	3	$R\,A_3$	$R\,B_3$	$Q\,A_2$	$Q\,B_2$	$P\,A_1$	$P\,B_1$
	4	$P\,A_1$	$P\,B_1$	$R\,A_3$	$R\,B_3$	$Q\,A_2$	$Q\,B_2$
	5	$Q\,A_2$	$Q\,B_2$	$P\,A_1$	$P\,B_1$	$R\,A_3$	$R\,B_3$
	6	$R\,A_3$	$R\,B_3$	$Q\,A_2$	$Q\,B_2$	$P\,A_1$	$P\,B_1$

crop-treatment sequences of such an experiment over the two preliminary years and six experimental years. Each of the sequences will be replicated three times, there being 9 blocks in all of two plots each. The yields for the six experimental years for crop P will consist of three replicates of Table 3, where a_1 , b_1 , a_2 , b_2 etc. represent the yields of

TABLE 3. YIELDS OF ONE REPLICATE OF CROP P FROM THE ROTATION EXPERIMENT OF TABLE 2

	Series					
	I		II		III	
Year	1	2	3	4	5	6
1	a_1	b_1				
2			a_2	b_2		
3					a_3	b_3
4	a_4	b_4				
5			a_5	b_5		
6					a_6	b_6

crop P in years 1, 2 etc. and in treatment cycles A and B (suffices being used to denote years). The results for the other two crops will be similar.

The analysis in this case is simple. The yields of crop P in series I will constitute a $2 \times 2 \times 3$ table of treatments $\times$ years (1 and 4) $\times$ replicates (blocks). The degrees of freedom of the analysis of variance of series I can therefore be partitioned as follows:

Treatments (T)	1
Years (Y)	1
Blocks (B)	2
$T \times Y$	1
$T \times B$	2
$Y \times B$	2
$T \times Y \times B$	2
	11

Since the treatments are repeated on the same plots the components T, B, and $T \times B$ which are derived from plot totals over the two years constitute the plot total part of the analysis, the remaining terms the plot $\times$ years part.

The contrasts between the totals of series I, II and III are estimates

of year differences (though subject to greater errors than the other components of years). The three single degrees of freedom for treatments from the three partial analyses combine into one for treatments and two for treatments × series (i.e. treatments × years). The combined analysis of the three series for one crop is therefore as shown in Table 4.

TABLE 4. ANALYSIS OF VARIANCE

		D.F.	M.S.	Expectation
	Series (Years)	2		
Plot Totals	Treatments (T)	1		
	Treatments × Series = $T \times Y$	2		
	Blocks (B)	6		
	$T \times B$ (plot error)	6	E_p	$2\sigma_p^2 + \sigma_w^2$
		15		
Plots × Years	Years (Y)	3		
	$T \times Y$	3		
	$Y \times B$	6		
	$T \times Y \times B$ (plot × year error)	6	E_w	σ_w^2
		18		
		35		

The $T \times B$ component with 6 d.f. provides an estimate of error (which we may term *plot error*) for comparisons based on plot totals over all years, and the $T \times Y \times B$ component, also with 6 d.f., provides an estimate (which we may term *plot × year error*) for comparisons not involving differences between plots. The estimate of the error variance of the difference between A and B averaged over all years, for example, will be $\frac{1}{6}E_p$, and that of the change in this difference between the first three years and the second three years will be $\frac{4}{9}E_w$.

If the errors of the plot yields are regarded as made up of two components, one independent from year to year with variance σ_w^2 and the other constant over all years with variance σ_p^2, the expectation of E_w is σ_w^2 and of E_p is $2\sigma_p^2 + \sigma_w^2$. (The coefficient of σ_p^2 is given by the number of yearly yields, two in this case, entering into each plot total.)

The sum of the sums of squares for the two estimates of error will be equal to the sum of the sums of squares for error in the analyses of the results of each year separately. There will be 2 d.f. for error in

each year giving 12 d.f. in all, and the expectation of each mean square will be $\sigma_p^2 + \sigma_w^2$, agreeing with the combined expectation of $T \times Y$ and $T \times Y \times B$.

If results are available for a larger multiple of three years the first part of the analysis will be unaltered except for change of divisors, but the degrees of freedom in the second part will be increased. With 12 years, for example, all these degrees of freedom will be multiplied by 3, and the expectation of E_p will be $4\ \sigma_p^2 + \sigma_w^2$. The analysis is also easily extended to an experiment containing more than two treatment sequences. With t treatment sequences a factor $t - 1$ will be introduced into the degrees of freedom of all components involving T. The alterations for change of rotation length are equally simple.

In experiments extending over a long period of time it is sometimes advisable, when trends are being considered, to subdivide the plots $\times$ years part of the analysis further, into say linear trend and remainder, so as to take account of possible differential trends of plots within blocks. This point is discussed by Cochran (1939) and Patterson (1953) and need not detain us here.

It will be noted that the interaction of years with blocks, $Y \times B$, has not been included in the estimate of the plot $\times$ year component of error. This is as it should be, since the different blocks may well exhibit year to year differences: such differences will not enter into the treatment comparisons since these are all compounded of differences within blocks in the same years. It is the elimination of this component which complicates the analysis of the rice-pasture experiment which is the subject of our investigation.

A rotation in which the same crop occurs more than once

If the same crop occurs more than once in a rotation comparisons can only be made with precision between the yields of the crop at the different stages of the rotation if the plots carrying them are arranged together in blocks. This requires that some at least of the different phases of the rotation must be grouped together in the same blocks. Whether all phases have to be so grouped depends on the crop rotation. In the four course rotation potatoes, wheat, potatoes, barley, for example, phases 1 and 3 must be grouped together, as must phases 2 and 4. In the rotation potatoes, potatoes, barley, wheat all four phases must be grouped together.

Whether comparisons between the same crop at different stages of a rotation are required will depend on the nature of the experiment. Such comparisons are usually of some interest, but may not be judged of sufficient value to compensate for the loss of accuracy resulting from

the larger blocks and the inconvenience and other troubles arising from the growing of more than one crop in the same block. We shall not discuss the possibilities here, as the problems are essentially those of design.

When the different phases of a rotation are grouped together new problems of analysis arise. We will first consider the simple example of a three-course rotation with two treatment cycles A and B as before, but with one crop repeated in two consecutive years e.g. with the rotation of crops P.P.R. instead of the rotation P.Q.R. of Table 1. Each replicate of the three phases of the rotation will be taken to be arranged together in a block instead of in three separate blocks. The yields of one replicate of crop P will now be as shown in Table 5.

TABLE 5. YIELDS OF THE DUPLICATED CROP IN ONE REPLICATE OF A THREE COURSE ROTATION WITH TWO CROPS THE SAME

	Sequence					
Year	1	2	3	4	5	6
1	a_1	b_1			a'_1	b'_1
2	a'_2	b'_2	a_2	b_2		
3			a'_3	b'_3	a_3	b_3
4	a_4	b_4			a'_4	b'_4
5	a'_5	b'_5	a_5	b_5		
6			a'_6	b'_6	a_6	b_6

Here a'_1 , b'_1 , etc. represent the yields of the second P crop under treatment cycles A and B. The contrast between $\frac{1}{2}$ $(\bar{a} + \bar{b})$ and $\frac{1}{2}(\bar{a}' + \bar{b}')$ is a *crop-phase* contrast. $\frac{1}{2}(\bar{a} - \bar{b} - \bar{a}' + \bar{b}')$ will then represent the treatment $\times$ crop-phase interaction.

The comparisons of chief interest will be between the means over all years of a, a', b and b'. The difference between $\frac{1}{2}(\bar{a} + \bar{a}')$ and $\frac{1}{2}(\bar{b} + \bar{b}')$ is derived from plot (i.e. sequence $\times$ replicate) totals over all years. The differences between $\bar{a}$ and $\bar{a}'$, and between $\bar{b}$ and $\bar{b}'$ do not involve plot differences.

If the same crop were grown in all years the interaction component of the (6 $\times$ 3) plot total $\times$ replicate table would represent plot error. In the present case, however, the plot totals over the six years will be affected by year differences. Sequences 1 and 2 contain years 1, 2, 4 and 5, sequences 3 and 4 years 2, 3, 5 and 6, etc. Hence the interaction component of the plot total $\times$ replicate table will contain components

of block × year interactions, which as pointed out above, require to be eliminated from error.

A partial analysis is easily effected. If plots 1 and 2, and the corresponding plots in the other replicates, are taken, we obtain a $2 \times 4 \times 3$ table of treatments × years × replicates, phases being completely confounded with years. The interaction of the (2×3) plot total × replicates table gives 2 d.f. for plot error, and the three factor interaction gives 6 d.f. for plot × year error. Similar analyses can be made for plots 3 and 4 and plots 5 and 6. Combining the results we obtain an estimate of plot error with 6 d.f. and an estimate of plot × year error with 18 d.f.

In experiments with a fair number of treatments this partial analysis may be judged sufficient to give adequate estimates of error, but it is of interest to consider how the remaining degrees of freedom for error can be recovered. In experiments containing rotations of varying length or with variations in rotations of the same length, without other treatments, as in the rice-pasture experiment under consideration, recovery of these degrees of freedom is essential.

The number of degrees of freedom for error from a single year's results will be 6. Hence there will be 36 d.f. in all for error. The full table of plot totals will consist of a 6×3 table of sequences × blocks. This indicates that the total number of degrees of freedom for plot error is $5 \times 2 = 10$. Hence there will be 26 d.f. for plot × year error. We therefore require an additional 4 d.f. for plot error and an additional 8 d.f. for plot × year error.

If c_1 is written for $a_1 + b_1$, c_1' for $a_1' + b_1'$ etc. the yields of one replicate can be written as shown in Table 6.

TABLE 6. COMBINED YIELDS OF ONE REPLICATE

	Sequences			
Year	1 and 2	3 and 4	5 and 6	Total
1	c_1	·	c_1'	Y_1
2	c_2'	c_2	·	Y_2
3	·	c_3'	c_3	Y_3
4	c_4	·	c_4'	Y_4
5	c_5'	c_5	·	Y_5
6	·	c_6'	c_6	Y_6
Total	$P_{1,2}$	$P_{3,4}$	$P_{5,6}$	R

Only the differences within columns of the c's and c''s, which have appeared as year differences, have so far been taken into account. We shall now consider the full analysis of this table.

There are 11 d.f. in all which can be partitioned into

Columns (plot pair totals)	2
Years	5
Columns × Years	4

Columns and years are not, however, orthogonal. To overcome this we may replace the column totals $P_{1,2}$ etc. by the quantities Q_1 , Q_2 , Q_3 , given by

$$\begin{aligned} Q_1 &= 2\{c_1 + c_2' + \tfrac{1}{2}(c_3 + c_3') + c_4 + c_5' + \tfrac{1}{2}(c_6 + c_6')\} \\ &= 2P_{1,2} - Y_{(1)} + R \end{aligned}$$

etc., where $Y_{(1)}$ represents the total $Y_1 + Y_2 + Y_4 + Y_5$ of the year × replicate totals for the years in which sequence 1 (or 2) carries the crop. All years are equally represented in these quantities, which are therefore orthogonal with years. They also represent differences between multiples of plot totals. Consequently if the quantities are calculated for all three replicates the interaction of the resultant 3 × 3 table of Q's will give 4 d.f. for plot error.

The divisor for the squares of the Q's, and the expectation of the error in terms of σ_p^2 and σ_w^2 , remain to be determined.

For the divisor we have

$$\begin{aligned} Q_1 - Q_2 &= 2\{c_1 - \tfrac{1}{2}(c_1 + c_1') + c_2' - c_2 + \tfrac{1}{2}(c_3 + c_3') - c_3' + \cdots\} \\ &= (c_1 - c_1') + 2(c_2' - c_2) + (c_3 - c_3') + \cdots \end{aligned}$$

Since each c is a total of two plots the divisor of $(Q_1 - Q_2)^2$ is $2\{1^2 \times 8 + 2^2 \times 4\} = 48$. The divisor of each Q^2 is therefore 24.

To evaluate the expectation in terms of σ_p^2 and σ_w^2 we replace all yields of plot 1 by p_1 , etc. We then have

$$Q_1 - Q_2 = 6p_1 + 6p_2 - 6p_3 - 6p_4$$

Hence the expectation of the σ_p^2 component of $(Q_1 - Q_2)^2$ is $4 \times 6^2 \times \sigma_p^2 = 144\sigma_p^2$. The expectation of the corresponding error mean square is therefore $3\sigma_p^2 + \sigma_w^2$. If years and plots had been orthogonal the expectation would have been $4\sigma_p^2 + \sigma_w^2$.

Alternatively the sum of the coefficients of the p^2 terms in $Q_1^2 + Q_2^2 + Q_3^2 - \frac{1}{3}(Q_1 + Q_2 + Q_3)^2$ can be calculated.

If these 4 d.f. for plot error and the 6 d.f. already obtained are combined by adding the sums of squares the expectation of the resultant mean square (10 d.f.) will be

$$\frac{4}{10}(3\sigma_p^2 + \sigma_w^2) + \frac{6}{10}(4\sigma_p^2 + \sigma_w^2) = 3.6\sigma_p^2 + \sigma_w^2$$

This method of combination is only fully efficient if σ_p^2 is small relative to σ_w^2. Otherwise greater weight should theoretically be given to the mean square with larger expectation. Even in the extreme case when σ_p^2 is very large relative to σ_w^2, however, the loss of information with the above weighting is quite trivial. Nevertheless, as will be seen later, the combined estimate is by no means equivalent to an estimate with 10 d.f. and expectation $4\sigma_p^2 + \sigma_w^2$.

The analysis of the three replicates of the c's can now be completed from the table of the totals of the c's over all replicates. It will take the form shown in Table 7, where the phase contrast $\bar{c} - \bar{c}'$ is represented by P.

TABLE 7. ANALYSIS OF VARIANCE OF TABLE 5

1. Years (ignoring treatments)	5	From the block
2. Blocks (replicates)	2	× year totals
3. $Y \times B$ (ignoring treatments)	10	table
4. Q's: totals over blocks ($P \times Y$)	2	As above
5. Q's: interaction with blocks (plot error)	4	
6. Remainder: totals over blocks (P, 1; $P \times Y$, 3)	4	By differences
7. Remainder: interaction with blocks (plot × year error)	8	
	35	

The sums of squares for the last two items are obtained by differences. The total sum of squares for the table of the totals of the c's over all replicates equals the total of the sums of squares for items 1, 4 and 6. The total of the sums of squares for all seven items is equal to the sum of squares for the 36 c's. Item 7 gives the remaining 8 d.f. for plot × year error.

It will be seen that the key to this analysis is given by the expressions Q which make the plot pair (column) totals orthogonal with years. In this example these expressions can be written down by inspection, but where this cannot be done they can be obtained by fitting constants for years (rows) and columns by the method of least squares. In the next section the method will be illustrated by application to the rice-pasture experiment.

The full analysis can now be completed. If required T, P, $T \times P$ and their interactions with years can be separately exhibited. It then takes the form shown in Table 8.

TABLE 8. FULL ANALYSIS OF VARIANCE

	D.F.	Expectation		D.F.	Expectation
T	1	$4\sigma_p^2 + \sigma_w^2$	Years (ignoring treatments)	5	
P	1	σ_w^2	Blocks	2	
$T \times P$	1	σ_w^2	$Y \times B$ (ignoring treatments)	10	
$T \times Y$	5	$0.4\sigma_p^2 + \sigma_w^2$	Plot error	10	$3.6\sigma_p^2 + \sigma_w^2$
$P \times Y$	5	$1.2\sigma_p^2 + \sigma_w^2$	Plot $\times$ years error	26	σ_w^2
$T \times P \times Y$	5	$1.2\sigma_p^2 + \sigma_w^2$		71	

T, P, $T \times P$, $T \times Y$ and $T \times P \times Y$ are computed in the ordinary manner. $P \times Y$ is obtained by subtraction of P from items 4 and 6 of Table 6. The errors of the various comparisons and the expectations of mean squares such as $T \times Y$ can be evaluated in terms of σ_p^2 and σ_w^2 by replacing yields by plot constants in the manner already indicated. The mean expectation of the 18 d.f. for treatments and treatments $\times$ years checks to $\sigma_p^2 + \sigma_w^2$, as it should.

It should be noted the plot error mean square no longer gives the error of differences based entirely on plot totals. Thus the error variance of the treatment difference $\frac{1}{2}(\bar{a} + \bar{a}') - \frac{1}{2}(\bar{b} + \bar{b}')$ is $\frac{2}{9}(4\sigma_p^2 + \sigma_w^2)$. The estimate of this will be $\frac{2}{9}(\frac{10}{9}E_p - \frac{1}{9}E_w)$. The accuracy of this estimate of error, relative to the direct estimate based on the 6 d.f. of the partial analysis is about equivalent to an extra 2 d.f. when σ_p^2 is small relative to σ_w^2, but equivalent to nearly an extra 4 d.f. when σ_p^2 is large relative to σ_w^2.

The analysis of the other crop of the rotation will be similar to the analysis of Table 4. There will now only be 2 d.f. for blocks, the remaining 4 d.f. representing year $\times$ block interactions. There will still only be 6 d.f. for plot $\times$ year error, since the other 4 d.f. will be completely confounded with year $\times$ block interactions.

Orthogonal partition of the degrees of freedom in the rice-pasture experiment.

In the rice-pasture experiment orthogonal Q functions for the 11 crop sequences cannot be written down by inspection. Fitting of constants must therefore be resorted to. The required functions can be obtained from the equations appropriate to a single replicate.

The yields and year and plot totals of one replicate are shown in Table 1. Let the year constants be $y_1, y_2 \cdots y_{12}$, and the plot constants $p_1, p_2, \cdots p_{11}$. The normal equations are then:

$$
\begin{aligned}
&4y_1 + p_1 + p_4 + p_7 + p_8 && = Y_1\\
&4y_2 + p_2 + p_4 + p_5 + p_9 && = Y_2\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&4p_1 + y_1 + y_4 + y_7 + y_{10} && = P_1\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&6p_4 + y_1 + y_2 + y_5 + y_6 + y_9 + y_{10} && = P_4\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&3p_8 + y_1 + y_5 + y_9 && = P_8\\
&\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

Denote the sum of the Y's for the years in which plot 1 carries rice by $Y_{(1)}$, so that $Y_{(1)} = Y_1 + Y_4 + Y_7 + Y_{10}$, etc. Further put $p_1 + p_2 + p_3 = p_{1-3}$, $p_4 + p_6 = p_{4,6}$, $p_5 + p_7 = p_{5,7}$, $p_8 + p_9 + p_{10} + p_{11} = p_{8-11}$, with similar totals for the P's.

Eliminating the y's in the p equation by means of the y equations we then have:

$$
\begin{aligned}
&12p_1 - 2p_{4,6} - 2p_{5,7} - p_{8-11} && = 4P_1 - Y_{(1)} && (1)\\
&12p_2 - 2p_{4,6} - 2p_{5,7} - p_{8-11} && = 4P_2 - Y_{(2)} && (2)\\
&12p_3 - 2p_{4,6} - 2p_{5,7} - p_{8-11} && = 4P_3 - Y_{(3)} && (3)\\
&18p_4 - 2p_{1-3} - 3p_{5,7} - 3p_8 - 3p_9 && = 4P_4 - Y_{(4)} && (4)\\
&18p_5 - 2p_{1-3} - 3p_{4,6} - 3p_9 - 3p_{10} && = 4P_5 - Y_{(5)} && (5)\\
&18p_6 - 2p_{1-3} - 3p_{5,7} - 3p_{10} - 3p_{11} && = 4P_6 - Y_{(6)} && (6)\\
&18p_7 - 2p_{1-3} - 3p_{4,6} - 3p_8 - 3p_{11} && = 4P_7 - Y_{(7)} && (7)\\
&9p_8 - p_{1-3} - 3p_4 - 3p_7 && = 4P_8 - Y_{(8)} && (8)\\
&9p_9 - p_{1-3} - 3p_4 - 3p_5 && = 4P_9 - Y_{(9)} && (9)\\
&9p_{10} - p_{1-3} - 3p_5 - 3p_4 && = 4P_{10} - Y_{(10)} && (10)\\
&9p_{11} - p_{1-3} - 3p_6 - 3p_7 && = 4P_{11} - Y_{(11)} && (11)
\end{aligned}
$$

We may now take the following sums of these equations:

(1) + (2) + (3)

$$12p_{1-3} - 6p_{4,6} - 6p_{5-7} - 3p_{8-11} = 4P_{1-3} - R$$

3(4) + 3(6) + (5) + (7)

$$36p_{4,6} - 16p_{1-3} - 12p_{8-11} = 12P_{4,6} + 4P_{5,7} - 4R$$

$$3(5) + 3(7) + (4) + (6)$$

$$36p_{5,7} - 16p_{1-3} - 12p_{8-11} = 12P_{5,7} + 4P_{4,6} - 4R$$

$$(8) + (9) + (10) + (11)$$

$$9p_{8-11} - 4p_{1-3} - 3p_{4,6} - 3p_{5,7} = 4P_{8-11} - R$$

It is easily verified that these equations are satisfied by

$$\left.\begin{aligned} p_{1-3} &= \tfrac{1}{4}P_{1-3} \\ p_{4,6} &= \tfrac{1}{6}P_{4,6} \\ p_{5,7} &= \tfrac{1}{6}P_{5,7} \\ p_{8-11} &= \tfrac{1}{3}P_{8-11} \end{aligned}\right\} \qquad (12)$$

This is a consequence of the fact that the sums of plots 1–3, 4 and 6, 5 and 7, and 8–11 are all orthogonal with years.

The values of the separate p's are then obtained by substitution in equations (1)–(11), and elimination of $p_8 - p_{11}$ from equations (4)–(7) by means of equations (8)–(11). This gives

$$\left.\begin{aligned} 12p_1 &= 4P_1 - Y_{(1)} - \tfrac{1}{3}P_{1-3} + \tfrac{1}{3}R \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \qquad (13)$$

$$\left.\begin{aligned} 24p_4 &= 5P_4 - P_6 + P_8 + P_9 - P_{10} - P_{11} - Y_{(4)} + Y_{(6)} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \qquad (14)$$

$$\left.\begin{aligned} 36p_8 &= 17P_8 - P_{10} + 3P_4 + 3P_7 - 5Y_{(8)} + Y_{(10)} + P_{1-3} \\ &\qquad - \tfrac{1}{2}P_{4,6} - \tfrac{1}{2}P_{5,7} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \qquad (15)$$

The total sum of squares accounted for by fitting the constants might be calculated from the expression $S(yY) + S(pP)$, the values of the year constants y being determined by substitution of the p values in the original normal equations, or alternatively from the expression $\frac{1}{4}S(Y^2) + S(pP')$, when $4P_1' = 4P_1 - Y_{(1)}$ etc. It is more satisfactory, however, to construct a set of orthogonal Q functions which partitions the sum of squares directly.

Such a set of functions can be constructed by taking appropriate groups of constants in turn in any desired order, and obtaining expressions for these in terms of the Y's and P's after eliminating all groups of constants already fitted. If certain conditions of symmetry are satisfied the sum of squares attributable to a group of constants will be calculable from the sum of the squares of the deviations of the values of the corresponding Q functions in the group.

As a first step we may replace the p's by $\bar{p}_{1-3}$, $\bar{p}_{4,6}$, $\bar{p}_{5,7}$ and $\bar{p}_{8-11}$, δp_1 , δp_2 , $\cdots$ δp_{11} , where $\bar{p}_{1-3} = \frac{1}{3}p_{1-3}$, $\bar{p}_{4,6} = \frac{1}{2}p_{4,6}$, etc. and $\delta p_1 = p_1 - \bar{p}_{1-3}$, $\delta p_4 = p_4 - \bar{p}_{4,6}$, etc. We may then take the following groups of constants in the order shown:

1. y_1 , y_2 , $\cdots$ y_{12}
2. $\bar{p}_{1-3}$, $\bar{p}_{4,6}$, $\bar{p}_{5,7}$, $\bar{p}_{8-11}$
3. δp_1 , δp_2 , δp_3
4. δp_8 , δp_9 , δp_{10} , δp_{11}
5. δp_4 , δp_6 and δp_5 , δp_7

The ordinary expressions of the analysis of variance give the sums of squares accounted for by groups 1 and 2, which are orthogonal.

The values of δp_1 , δp_2 , δp_3 before fitting δp_4 to δp_{11} are given by replacing p_4 by $\bar{p}_{4,6}$ etc. in equations (1)–(3). It will be seen that the equations (13) already obtained are unaltered, and the corresponding sum of squares will be given by

$$\frac{1}{\lambda}\mathrm{dev}^2\{Q_1 , Q_2 , Q_3\}$$

where $Q_1 = 4P_1 - Y_{(1)}$ etc., dev^2 indicates the sum of the squares of the deviations, and λ is a divisor to be determined.

We have, by ordinary least squares procedure,

$$V(\delta t_1 - \delta t_2) = (c_{11} - 2c_{12} + c_{22})\sigma^2$$

where c_{11} and c_{12} are the values of δt_1 and δt_2 when $P_1 = +\frac{2}{3}$, $P_2 = P_3 = -\frac{1}{3}$, $P_{4,6} = P_{5,7} = P_{8-11} = 0$ and all $Y = 0$, and c_{21} $(= c_{12})$ and c_{22} are the values of δt_1 and δt_2 when $P_2 = \frac{2}{3}$, $P_1 = P_3 = -\frac{1}{3}$, $P_{4,6} = P_{5,7} = P_{8-11} = 0$, and all $Y = 0$. From equations (13)

$$12\ \delta p_1 = 4P_1 - Y_{(1)} - \frac{4}{3}P_{1\text{-}3} + \frac{1}{3}R$$

Thus

$$c_{11} = c_{22} = \frac{1}{12}\cdot\frac{8}{3} = \frac{2}{9}$$

$$c_{12} = -\frac{1}{12}\cdot\frac{4}{3} = -\frac{1}{9}$$

and

$$c_{11} - 2c_{12} + c_{22} = \frac{2}{3}$$

Hence

$$V(\delta t_1 - \delta t_2) = V\frac{1}{12}(Q_1 - Q_2) = \frac{2}{3}\sigma^2$$

Hence

$$\lambda = \frac{1}{2}\cdot 12^2 \cdot \frac{2}{3} = 48.$$

The values for δp_8 to δp_{11} before fitting δp_4 to δp_7 are given by equations (8)–(11), replacing p_4 and p_6 by $\bar{p}_{4,6}$ and p_5 and p_7 by $\bar{p}_{5,7}$. Hence

$$9p_8 = 4P_8 - Y_{(8)} + \tfrac{1}{4}P_{1\text{-}3} + \tfrac{1}{4}P_{4,6} + \tfrac{1}{4}P_{5,7}$$

By putting $P_8 = +\frac{3}{4}$, $P_9 = P_{10} = P_{11} = -\frac{1}{4}$ and all other P and all $Y = 0$ we obtain $c_{88} = +\frac{1}{3}$, $c_{89} = -\frac{1}{9}$, and therefore $c_{88} - 2c_{89} + c_{99} = \frac{8}{9}$. Hence the divisor is 36.

Finally the values for δp_4 to δp_7 after fitting all other constants are given by equations (14). We have

$$12(\delta p_4 - \delta p_6) = 3(P_4 - P_6) + P_8 + P_9 - P_{10} - P_{11} - Y_{(8)} - Y_{(9)} + Y_{(10)} + Y_{(11)}$$

By putting $P_4 = +\frac{1}{2}$, $P_6 = -\frac{1}{2}$ and all other P and all $Y = 0$ we obtain

$$c_{44} - c_{46} = \tfrac{1}{4}$$

and therefore $c_{44} - 2c_{46} + c_{66} = \frac{1}{2}$. Hence the divisor of 144 $(\delta p_4 - \delta p_6)^2$ is 72.

We may therefore summarize the analysis of a single replicate as follows:

In addition to the year totals, $Y_1 - Y_{12}$, and the totals of the four groups of plots orthogonal with years P_{1-3}, $P_{4,6}$, $P_{5,7}$ P_{8-11}, calculate the following 11 quantities $Q_1 - Q_{11}$.

For plots 1–3: $Q_1 = 4P_1 - Y_{(1)}$, etc. Total: $4P_{1-3} - R$

" " 4–7: $Q_4 = 3P_4 + P_8 + P_9 - Y_{(8)} - Y_{(9)}$ etc.

Total: $3P_{4,6} + 3P_{5,7} + 2P_{8-11} - 2R$

" " 8–11: $Q_8 = 4P_8 - Y_{(8)}$, etc. Total: $4P_{8-11} - R$

The expressions for the sums of squares are shown in Table 9.

The reason why the groups were taken in the chosen order will now be clear. If $\delta p_8 - \delta p_{11}$ had been taken after $\delta p_4 - \delta p_7$ equations (15) would have been used for determining c_{88}, c_{89}, etc. In this case $c_{8,9}$ would not be equal to $c_{8,10}$, etc., i.e. not all the covariances between $\delta p_8 - \delta p_{11}$ would be equal. Consequently the expression for the sum

TABLE 9. ANALYSIS OF VARIANCE: EXPRESSIONS FOR SUMS OF SQUARES

		D.F.	S.S.
Years		11	$\frac{1}{4}\,\mathrm{dev}^2(Y_1\ ,\ Y_2\ ,\ \cdots\ Y_{12})$
Plot Totals	Orthogonal groups	3	$\frac{1}{12}\,\mathrm{dev}^2(P_{1\text{-}3}\ ,P_{4,6}\ ,P_{5,7}\ ,P_{8\text{-}11})$
	Within groups 1–3	2	$\frac{1}{48}\,\mathrm{dev}^2(Q_1\ ,\ Q_2\ ,\ Q_3)$
	" groups 4, 6 and 5, 7	2	$\frac{1}{72}\,\{(Q_4 - Q_6)^2 + (Q_5 - Q_7)^2\}$
	" group 8–11	3	$\frac{1}{36}\,\mathrm{dev}^2(Q_8\ ,\ Q_9\ ,\ Q_{10}\ ,\ Q_{11})$
	Total	10	
Remainder (plots × years)		26	
		47	

of squares for $\delta p_8 - \delta p_{11}$ would be more complicated, without any compensating simplification in the sum of squares for $\delta p_4 - \delta p_7$.

The divisors in the above sums of squares may be checked by setting out the coefficients of the differences of pairs of constants and summing their squares. Thus $9(\delta p_8 - \delta p_9)$ has the coefficients shown in Table 10, with sum of squares 72.

TABLE 10. COEFFICIENTS OF $9(\delta p_8 - \delta p_9)$

Year	1	2	3	4	5	6	7	8	9	10	11	Total
1	−1	·	·	−1	·	·	−1	+3	·	·	·	0
2	·	+1	·	+1	+1	·	·	·	−3	·	·	0
5	·	−1	·	−1	·	·	−1	+3	·	·	·	0
6	·	·	+1	+1	+1	·	·	·	−3	·	·	0
9	·	·	−1	−1	·	·	−1	+3	·	·	·	0
10	+1	·	·	+1	+1	·	·	·	−3	·	·	0
Total	0	0	0	0	+3	0	−3	+9	−9	0	0	0

Similarly the orthogonality of the different components can be checked by verifying that the sum of the products of the corresponding coefficients of the differences of each pair of constants is zero.

An alternative numerical check of the divisors and of orthogonality can be made by assuming values of the year and plot constants, building up the plot yields from them, and carrying out the analysis. If the

expressions are correct the residual sum of squares (26 d.f.) should be zero. This provides a useful overall check and a test of the numerical procedure, though it should be remembered that a single numerical test does not necessarily detect all errors. Taking $y_1 = 1$, $y_2 = 2$, $\cdots$ $y_{12} = 12$, and $p_1 = 1$, $p_2 = 2$, $p_{11} = 11$, the yields of plot 1 are 2, 5, 8, 11, those of plot 2 are 4, 7, 10, 13, etc. and the analysis shown in Table 11 is obtained.

TABLE 11. NUMERICAL CHECK

	D.F.	S.S.
Years	11	659.75
Orthogonal groups	3	344.25
Q_1 , Q_2 , Q_3	2	6
$Q_4 - Q_6$, $Q_5 - Q_7$	2	16
Q_8 , Q_9 , Q_{10} , Q_{11}	3	7.25
	21	1033.25

The total agrees with the total sum of squares of deviations of all yields (47 d.f.)

Scheme of analysis of the rice-pasture experiment

We may now complete the scheme of analysis for the rice-pasture experiment. Table 12 shows the general scheme, which has the same pattern as Table 8.

TABLE 12. ANALYSIS OF VARIANCE OF THE RICE-PASTURE EXPERIMENT

		D.F.
Plot Totals	Crop sequences	10
	Blocks	2
	Plot error	20
	Total	32
	Years	11
Plots × Years	Crop sequences × years	26
	Blocks × years	22
	Plot × year error	52
	Total	100
		143

Tables corresponding to Table 1 must be prepared for each replicate, and for the total of all three replicates. The sums of squares for blocks, years and blocks × years are computed in the ordinary manner from the marginal totals for years. For crop sequences and plot error the quantities P_{1-3} , $P_{4,6}$, $P_{5,7}$, P_{8-11} and $Q_1 - Q_{11}$ must be calculated for each replicate and for the total of all replicates. These are set out in four tables, namely that for the P totals and those for Q_1 to Q_3 , Q_4 to Q_7 , and Q_8 to Q_{11} . Columns of the differences $Q_4 - Q_6$ and $Q_5 - Q_7$ are also required. The marginal totals for all replicates (with an additional factor 3 in the divisors) give the components of the sum of squares for crop sequences (10 d.f.) while the interaction components (which for $Q_4 - Q_6$ and $Q_5 - Q_7$ are merely the sums of the squares of the deviations) give the components of the sum of squares for plot error (20 d.f.).

The sum of squares for crop sequences × years (26 d.f.) is then obtained by subtracting the sums of squares for years and for crop sequences from the total sum of squares (47 d.f.) for the crop sequence × years table (totals over all replicates). Likewise the sum of squares for plot × year error (52 d.f.) is obtained by subtracting all other items from the sum of squares (143 d.f.) for the whole experiment.

It remains to determine the expectations of the various components of the plot error sums of squares in terms of σ_p^2 and σ_w^2 . The procedure already explained of evaluating the sums of squares of the coefficients of the plot constants is followed. The group 8–11 component can be broken down into 3 orthogonal components. We may take the orthogonal components $9(\delta p_8 - \delta p_9)$, $9(\delta p_{10} - \delta p_{11})$ and $9(\delta p_8 + \delta p_9 -$

TABLE 13. EXPECTATIONS OF PLOT-ERROR MEAN SQUARES

		D.F.	Expectation
Orthogonal groups	$P_{4,6} - P_{5,7}$	2	$6\sigma_p^2 + \sigma_w^2$
	$P_{1,3} - P_{8,11}$	2	$\frac{7}{2}\sigma_p^2 + \sigma_w^2$
	$P_{4,6} + P_{5,7} - P_{1,3} - P_{8,11}$	2	$\frac{19}{4}\sigma_p^2 + \sigma_w^2$
Within group 1–3		4	$3\sigma_p^2 + \sigma_w^2$
" groups 4, 6 and 5,7		4	$4\sigma_p^2 + \sigma_w^2$
" group 8–11	$\delta p_8 - \delta p_9$ and $\delta p_{10} - \delta p_{11}$	4	$\frac{5}{2}\sigma_p^2 + \sigma_w^2$
	$\delta p_8 + \delta p_9 - \delta p_{10} - \delta p_{11}$	2	$\frac{11}{4}\sigma_p^2 + \sigma_w^2$
	Total	20	$3.6\sigma_p^2 + \sigma_w^2$

$\delta p_{10} - \delta p_{11}$). The σ_p^2 component of $81(\delta p_8 - \delta p_9)^2$, for example, obtained from the total line of Table 8, is $(3^2 + 3^2 + 9^2 + 9^2)\ \sigma_p^2 = 180\ \sigma_p^2$. Since the divisor is 72 the expectation of the corresponding sum of squares (1 d.f.) is $\frac{5}{2}\sigma_p^2 + \sigma_w^2$. A similar procedure can be followed for the orthogonal groups component. The complete set of expectations is shown in Table 13.

The expectation for the whole 72 d.f. for plot and plot × year error is therefore

	D.F.	M.S.	Expectation
Plot error	20	E_p	$3.6\ \sigma_p^2 + \sigma_w^2$
Plot × year error	52	E_w	σ_w^2
Total	72		$\sigma_p^2 + \sigma_w^2$

This is as it should be, since the 72 d.f. is made up of 6 d.f. for error for each year separately, each with expectation $\sigma_p^2 + \sigma_w^2$. The expectation for plot error could of course be obtained by utilising this fact, but direct evaluation provides a useful check and exhibits the structure of the analysis.

In the above analysis the degrees of freedom for crop sequences and for crop sequences × years have been partitioned in a manner suitable for the separation of plot error and plot × year error. They can also be partitioned so as to isolate the various contrasts which are of experimental interest. The main contrasts are those between rotations A and C and the two phases B and B' of B. The contrasts between A and C and the mean of B and B' are part of the 10 d.f. for crop sequences. The sum of squares (2 d.f.) is given by

$$\frac{1}{36}T_{1\text{-}3}^2 + \frac{1}{72}T_{4\text{-}7}^2 + \frac{1}{36}T_{8\text{-}11}^2 - \frac{1}{144}T^2$$

where T_{1-3} represents the total over the three replicates of P_{1-3} , etc., and T the grand total. The corresponding mean square is not, however, directly comparable with the plot error mean square, since its expectation, from the results already given, is $\frac{1}{2}(\frac{7}{2} + \frac{19}{4})\ \sigma_p^2 + \sigma_w^2 = \frac{33}{8}\ \sigma_p^2 + \sigma_w^2$. It must therefore be compared with $\frac{55}{48}\ E_p - \frac{7}{48}\ E_w$.

The contrast between B and B' is wholly within plots and is therefore subject to plot × year error, the error variance being $\frac{1}{12}\ E_w$, and the sum of squares (1 d.f. from crop sequences and years) $\frac{1}{12}\{S(B) - S(B')\}^2$.

The remaining 33 d.f. from crop sequences and crop sequences × years represent the interactions of rotations A and C and the two phases of rotation B with years. The expectation of the corresponding mean square is $\frac{1}{33}(36 - \frac{33}{4})\sigma_p^2 + \sigma_w^2 = \frac{37}{44}\sigma_p^2 + \sigma_w^2$ and it is therefore compared with $\frac{185}{792}E_p + \frac{607}{792}E_w$.

It should also be noted that blocks × years contains components of plot differences.

Analysis of incomplete data

In long term agricultural experiments an interim analysis of the results is often required. The interim data usually lack the balance that is attained when the experiment has run its full term. In the type of experiment we have just been discussing an exact analysis may then prove excessively laborious since the relevant normal equations will no longer be readily soluble. On the other hand it is unreasonable to tell the experimenter that he must await the completion of the experiment before attempting to draw any conclusions from the results. The statistician must therefore be prepared to determine how far approximate methods of analysis will enable interim reports to be made on long-term experiments in an expeditious manner and without undue labour.

In the case of the rice-pasture experiment there do not appear to be any approximate methods which are very satisfactory. As an example of the problems involved we may consider the analysis of 6 years' data. We may adopt the same notation as previously except that P_1 etc. and T_1 etc. now represent totals over 6 years instead of 12. In the exact analysis on the lines already laid down, the degrees of freedom would partition in the manner shown in Table 14, which corresponds to Table 9.

TABLE 14. ANALYSIS OF VARIANCE OF 6 YEARS' DATA

		D.F.
Plot Totals	Crop sequences	10
	Blocks	2
	Plot error	20
	Years	5
Plots × years	Crop sequences × years	8
	Blocks × years	10
	Plot × year error	16
		71

An estimate of plot error could be built up from the orthogonal groups and the Q functions already obtained. The components derived from the Q functions are now, however, no longer fully orthogonal amongst themselves (though they are still, as is essential, orthogonal with years). This will not seriously affect the estimate of plot error—it has an analogous effect to the inclusion of some degrees of freedom more than once in an estimate of error—but the sum of squares so obtained cannot now be subtracted from the total sum of squares for error to give the plot × year error without serious risk of disturbance due to non-orthogonality.

The simplest procedure, therefore, is to throw the block × year interaction into the estimates of error. The interaction of the table of treatment sequence totals × blocks (appropriate allowance being made for differing numbers of yields in the totals) will then give an estimate of plot error (containing components of block × year interaction) with 20 d.f. The total of the interaction components of the 11 tables of treatment sequences × replicates (the last two of which contribute nothing) will give an estimate of plot × year error (also containing block × year interaction) with 26 d.f. The expectation in terms of σ_p^2 and σ_w^2 of the plot error can be evaluated by the procedure already adopted.

This method of estimation of error will give estimates which are too large if block × year interactions are substantial. This point can be examined by comparing the mean square for block × year interactions, calculated from the table of block × year totals, with its error expectation in terms of σ_p^2 and σ_w^2 . If it is considered that the interaction should be eliminated then a full least square analysis, with inversion of the matrix, will have to be undertaken. Construction of orthogonal functions, which will necessarily be complicated, will not be worth while.

REFERENCES

Cochran, W. G. 1939. Long-term agricultural experiments. *J. R. Statist. Soc. Suppl., 6*, 104–148.

Crowther, F. and Cochran, W. G. 1942. Rotation experiments with cotton in the Sudan Gezira. *J. Agric. Sci., 32*, 390–405.

Patterson, H. D. 1953. The analysis of the results of a rotation experiment on the use of straw and fertilizers. *J. Agric. Sci., 43*, 77–88.

Yates, F. 1949. The design of rotation experiments. *Comm. Bur. Soil Sci. T. C.* No. 46

THE DERIVATION OF JOINT DISTRIBUTION AND CORRELATION BETWEEN RELATIVES BY THE USE OF STOCHASTIC MATRICES

C. C. Li and Louis Sacks

Department of Biostatistics, Graduate School of Public Health, and Department of Mathematics, University of Pittsburgh, Pittsburgh, Pa.

The absolute frequencies of the various genotypic parent-offspring and sib-pair combinations with respect to one pair of genes, autosomal or sex-linked, are well known to geneticists, for they are of practical value in certain types of studies in human heredity. These frequencies are, however, obtained by a rather long process, even for the simple case of full sibs. The procedure of obtaining the frequencies of uncle-nephew or first cousin combinations is entirely too tedious even with the help of matrix notations (Hogben, 1933). The purpose of the present communication is three-fold. The first is to give a simple procedure of finding the frequencies of the various genotype combinations of near relatives by using matrices of *conditional probabilities*. The second purpose is to express such matrices of conditional probabilities of relatives in the form of a linear function of some basic matrices. The third is to deduce the correlations between relatives from such linear functions of the basic matrices. The meaning of these statements will be made clear in later sections.

For many years there existed two apparently very different methods of obtaining the genotypic correlations between relatives. One is a straightforward but long procedure by which the frequencies of the various combinations of the stated relatives in the general population are first found and then the correlation is calculated from such a "correlation table". On the other hand, they may be obtained by the method of path coefficients, developed by Wright (1921 and later). The latter method gives the required correlation coefficient almost instantly once the relationship is specified, but does not give us any information about the frequencies of the various combinations of the relatives in the population. Moreover, one has to be familiar with the mathematical theorems concerning the path coefficients before he can use them to derive the required correlations. The method of stochastic process and its final reduction to some basic matrices, as described in the following pages, will give us both the frequencies of relative-pairs and their correlation in a very simple manner. It is hoped that the present method will bridge the gap between the two existing procedures.

AUTOSOMAL GENES

1. Instead of directly deriving the absolute joint distribution with regard to the genotypes of certain relative pairs (e.g. parent-offspring, uncle-nephew, first-cousins, etc.), we shall deal with the *conditional* probabilities that a relative should be of a certain genotype when the other relative's genotype is given. Thus, when an uncle is given to be AA, we are to find the probabilities that his nephew should be AA, Aa, aa, repectively. These three probabilities will of course add up to unity. Similar probabilities may be found for Aa and aa uncles. Thus, we obtain the following array of probabilities (t):

		Nephew		
		AA	Aa	aa
Condition:	Uncle AA	t_{22}	t_{21}	t_{20}
Condition:	Uncle Aa	t_{12}	t_{11}	t_{10}
Condition:	Uncle aa	t_{02}	t_{01}	t_{00}

where the sum of the three probabilities in each row is unity. Such an array of probabilities is known as the matrix of "transition probabilities", from uncle to nephew in this case. Once such a matrix is obtained, the absolute frequencies of the various genotypic combinations of uncle and nephew in the general random mating population may be obtained immediately by multiplying the uncle conditions by their respective "initial" probabilities; that is, multiplying the first row by p^2, the second row by $2pq$ and the third row by q^2, where p is the frequency of gene A in the population ($q = 1 - p$). Therefore, we shall largely deal with the transition matrices of relatives in the following.

2. In order to facilitate the derivation of transition matrices for various types of relatives it is convenient to introduce another consideration besides the genotypes. Let X be a gene of the locus under consideration. It may be A or a. We say that it exists in two *states*. If we select a gene from each of two unrelated individuals and they happen to be both A or both a, we say that these two genes are *alike in state*. On the other hand, if a parent has an X-gene which is transmitted to his child, we say that these two X-genes (one drived from the other) are *identical by descent*. Genes of the latter kind are necessarily alike in state, barring mutation, but the reverse is not true. The distinction between the two ways in which two alleles may be alike has been pointed out by a number of geneticists (for example, Cotterman 1941; Malécot 1948; Crow 1954). To trace the origin of a gene, we may label each independent gene by a subscript (a serial number for identification). Thus, in a random mating population, the two parents may be represented by $X_1X_2 \times X_3X_4$, whatever their genotypes.

3. A pair of relatives may have both, one, or none genes identical by descent, depending upon the type of their relationship. The matrices of transition probabilities (with regard to genotypes) for the three cases are respectively:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad T = \begin{bmatrix} p & q & 0 \\ \frac{1}{2}p & \frac{1}{2} & \frac{1}{2}q \\ 0 & p & q \end{bmatrix}, \quad 0 = \begin{bmatrix} p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \end{bmatrix}.$$

The truth of I and 0 is obvious. If two relatives share both genes identical by descent, they are necessarily of the same genotype. On the other hand, if they share no gene identical by descent, the probabilities that the other relative's being AA, Aa, aa are p^2, $2pq$, q^2, whatever the genotype of the given relative. The truth of T is also easily seen. For instance, if one relative is given to be AA and the other relative shares one gene identical by descent, then the probability that the latter should be Aa is q, which is the probability that the other (non-identical) gene is a.

4. With the above three basic transition matrices, it is easy to derive the conditional probabilities for any specified relatives. Let us consider first some of the simple *unilineal* relatives. Now, a pair of parent and offspring necessarily share one gene identical by descent. Therefore, the transition probabilities from parent to offspring are simply the elements of T. This may be readily verified by the long classical method of considering all the possible mating types in the population. It should be noted that T also gives the transition probabilities from offspring to parent.

Next, consider the transition probabilities from grandparent to grandchild. If the grandparent is given to be AA, the probabilities of the genotypes of the parent are the elements of the first row of T. For each genotype of the parent, the probabilities that the child should be AA are the elements of the first column of T. Thus, the total conditional probability that the grandchild should be AA is the sum of the products of the corresponding elements of the first row and of the first column of T; viz.

$$(p, q, 0) \begin{bmatrix} p \\ \frac{1}{2}p \\ 0 \end{bmatrix} = p^2 + \tfrac{1}{2}pq = \tfrac{1}{2}p + \tfrac{1}{2}p^2.$$

Similar considerations show that the conditional probabilities for the various genotypes of grandchild with *one* grandparent specified are

given by the elements of the product matrix $T \cdot T = T^2$. Labeling and tracing the origin of genes shows that a grandparent-grandchild pair has an equal chance to share one gene identical by descent and none at all. Therefore,

$$T^2 = \tfrac{1}{2}T + \tfrac{1}{2}0$$

as may be readily verified by direct multiplication of T by T. Since T represents the transition matrix from child to parent as well, T^2 also gives the conditional probabilities for half-sibs, who have one parent in common. The above relation shows that the tedious multiplication of matrices is reduced to the simple operation of addition. Further investigation shows that,

$$T^3 = \tfrac{1}{4}T + \tfrac{3}{4}0; \qquad T^4 = \tfrac{1}{8}T + \tfrac{7}{8}0,$$

and in general,

$$T^{n+1} = (\tfrac{1}{2})^n T + (1 - (\tfrac{1}{2})^n)0$$

where $n + 1$ is the number of generations between the two relatives.

5. Now, we may turn to *bilineal* relatives, the most important type of which is full sibs. Let the given sib be X_1X_3. Since X_1 comes from one parent and X_3 comes from the other, his parents may be designated by $X_1X_2 \times X_3X_4$, in whatever states the genes may exist. Then the probability that the other sib should also receive X_1 from the first parent and X_3 from the second parent is $\frac{1}{4}$. That is to say, the probability that the other sib should have both genes identical by descent is $\frac{1}{4}$. This probability is independent of the genotypes of the sibs, independent of the genotypes of their parents and independent of the gene frequencies of the general population. Therefore, the three kinds of sib-pairs, *AA–AA*, *Aa–Aa*, *aa–aa*, should have a component of $\frac{1}{4}$ in all populations. In fact, such sib-pairs are equivalent to "identical twins" as far as the *A–a* locus is concerned.

On the other hand, the probability that the other sib should be X_2X_4, sharing no gene identical by descent with the given X_1X_3 sib, is also $\frac{1}{4}$. Finally, the probability that the sibs share one gene (X_1 or X_3) identical by descent is $\frac{1}{2}$. Hence the transition probabilities for full sibs are the elements of the matrix.

$$S = \tfrac{1}{4}I + \tfrac{1}{2}T + \tfrac{1}{4}0.$$

This, again, may be verified by the long classical method of considering all possible matings in the general population.

Another relationship in which the relatives can share both genes identical by descent is that of double first cousins, whose parents are

members of two sibships. There are six types of sibships in the population, corresponding to the six types of matings. Double cousinship can occur in twenty-one distinct ways, since any pair of the six types of sibship may be taken. The absolute frequencies of the various double cousin pairs in the population may be calculated by random mating such pairs of sibships. This is the procedure employed by Fisher (1918) and it involves a large amount of algebra. The conditional frequencies of double cousins, however, may be obtained in the following simple manner.

We label and trace the eight independent genes of the four grandparents involved in the two families, as illustrated in Fig. 1. When one (any one) cousin is given, the probabilities that the other cousin should have both, one, or none genes identical by descent are $\frac{1}{16}$, $\frac{6}{16}$, $\frac{9}{16}$, respectively. Hence the matrix of conditional probabilities for double

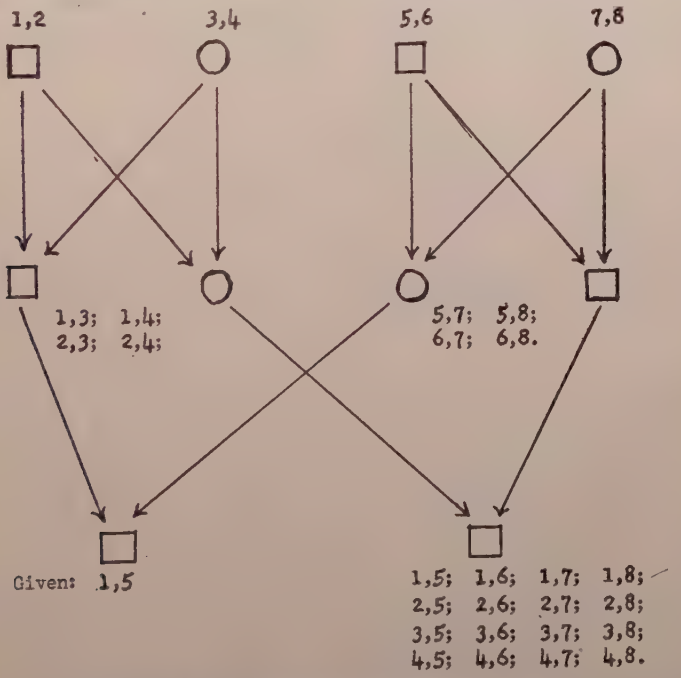


FIG. 1. RELATIONSHIP BETWEEN DOUBLE FIRST COUSINS. THE NUMERALS 1, . . . , 8 ARE LABELS OF THE EIGHT GENES OF THE FOUR GRANDPARENTS INVOLVED IN THE TWO FAMILIES. THE FOUR PAIRS OF LABELS BELOW THE PARENTS INDICATE THEIR POSSIBLE GENETIC CONSTITUTION ACCORDING TO ORIGIN OF GENES, ETC. FOR EACH GIVEN COUSIN, THERE ARE 16 POSSIBILITIES FOR THE OTHER COUSIN, OF WHICH ONE IS IDENTICAL (1, 5), THREE WITH $GENE_1$ BUT WITHOUT $GENE_5$, THREE WITH $GENE_5$ BUT WITHOUT $GENE_1$—A TOTAL OF SIX POSSIBILITIES SHARING ONE IDENTICAL GENE. THE REMAINING NINE GENOTYPES ARE INDEPENDENT OF THE GIVEN CONDITION.

first cousins is,

$$S^2 = \frac{1}{16} I + \frac{6}{16} T + \frac{9}{16} 0.$$

This matrix, it is to be noted, happens to be the "square" of the S matrix for siblings.

6. Now, we shall consider the unilineal relatives for whom both S and T matrices are involved. The term "uncle-nephew" is used here in its broad sense, including uncle-niece, aunt-nephew and aunt-neice relations as we are dealing with autosomal genes. When an uncle is specified, the conditional probabilities for his nephew (his full sib's child) will be given by the product matrix ST. Conversely, when a nephew is given, the conditional probabilities for his uncle (his parent's full sib) will be given by the product matrix TS. It may be easily verified that $ST = TS$, showing that the uncle-nephew transition matrix is also the nephew-uncle matrix. But, the most remarkable property is that

$$ST = TS = T^2$$

indicating that uncle-nephew relations are the same as those for grandparent-grandchild or half sibs, as established by other methods in human genetics.

Similarly, one's first cousin is his parent's full sib's child; hence the conditional probabilities for first cousins are the elements of the matrix

$$TST = T^3$$

which also gives the conditional probabilities for the great-grandchild when one great-grandparent is given. In a similar manner other relatives may be expressed by a product matrix involving T and S.

In the light of the results derived above, we see that Feller's conclusions (1950, pp. 102–103) that "brothers are as close to each other as grandfather and grandson" (equivalent to saying $S = T^2$) and that the conditional probabilities for a great-grandson or a nephew when a great-grandfather or an uncle is given are the same (equivalent to saying $ST = T^3$) are both incorrect. These errors are due to his failure to use the sibling transition matrix S in his calculations.

SEX-LINKED GENES

7. The above method of conditional probabilities may be applied to sex-linked characters with only slight modifications. Without loss of generality, we let the heterogametic sex be males (A or a) and the

homogametic sex be females (AA or Aa or aa) as in the case of mankind. There will be four parent-offspring transition matrices (T) according as the parent and offspring are males or females. To distinguish them, we shall use two subscripts, the first to denote the number of rows and the second the number of columns in the matrix of transition probabilities. Thus, T_{33} , T_{32} , T_{23} , T_{22} , are the matrices for mother-daughter, mother-son, father-daughter, father-son, respectively. The matrix T_{33} is the same as T for autosomal genes. The remaining three are as follows:

$$T_{32} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}; \qquad T_{23} = \begin{bmatrix} p & q & 0 \\ 0 & p & q \end{bmatrix}, \qquad T_{22} = \begin{bmatrix} p & q \\ p & q \end{bmatrix}$$

As before, 0 denotes an independence matrix, with appropriate subscripts. Thus,

$$0_{32} = \begin{bmatrix} p & q \\ p & q \\ p & q \end{bmatrix}, \qquad 0_{23} = \begin{bmatrix} p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \end{bmatrix}$$

In this notation, the father-son transition matrix $T_{22} = 0_{22}$, indicating that the genotypes of father and son are independent. Similar to the T's, there are four sibling transition matrices, according as the given sib is a male or female. The following are for sister-sister, sister-brother, brother-sister, brother-brother, respectively (subscript 3 = sister, 2= brother):

$$S_{33} = \tfrac{1}{2}I_{33} + \tfrac{1}{2}T_{33}\,, \qquad S_{32} = \tfrac{1}{2}T_{32} + \tfrac{1}{2}0_{32}\,,$$

$$S_{23} = \tfrac{1}{2}T_{23} + \tfrac{1}{2}0_{23}\,, \qquad S_{22} = \tfrac{1}{2}I_{22} + \tfrac{1}{2}0_{22}\,.$$

The coefficients in the linear expression of each S give us the nature of the siblings. For example, $S_{22} = \frac{1}{2}I_{22} + \frac{1}{2}0_{22}$ means that the two brothers either have the same identical gene or have two independent genes.

8. The matrix of conditional probabilities for other relatives may be obtained by multiplying the appropriate matrices, just as in the case of autosomal genes. But the kinds of relatives for sex-linked genes are many times more than the corresponding relatives for autosomal genes as each individual involved has to be distinguished in regard to sex. It is unnecessary to enumerate all the near relatives. We shall illustrate

the method by considering grandparent-grandchild and relatives of the "uncle-nephew" type. There are eight different kinds of relations for each type. The results are given in Table 1 in which we assume that the elder relative is being given.

Note from the upper part of Table 1 that whenever two males are

TABLE 1. MATRIX OF CONDITIONAL PROBABILITIES FOR SEX-LINKED GENES

When the grandparent is given

	grandmother-granddaughter	grandfather-granddaughter	grandmother-grandson	grandfather-grandson
maternal	$T_{33}T_{33}=\frac{1}{2}T_{33}+\frac{1}{2}0_{33}$	$T_{23}T_{33}=\frac{1}{2}T_{23}+\frac{1}{2}0_{23}$	$T_{33}T_{32}=\frac{1}{2}T_{32}+\frac{1}{2}0_{32}$	$T_{23}T_{32}=\frac{1}{2}I_{22}+\frac{1}{2}0_{22}$
paternal	$T_{32}T_{23}=T_{33}$	$T_{22}T_{23}=0_{23}$	$T_{32}T_{22}=0_{32}$	$T_{22}T_{22}=0_{22}$

When the Aunt or Uncle is given

	aunt-niece	uncle-niece	aunt-nephew	uncle-nephew
maternal	$S_{33}T_{33}=\frac{3}{4}T_{33}+\frac{1}{4}0_{33}$	$S_{23}T_{33}=\frac{1}{4}T_{23}+\frac{3}{4}0_{23}$	$S_{33}T_{32}=\frac{3}{4}T_{32}+\frac{1}{4}0_{32}$	$S_{23}T_{32}=\frac{1}{4}I_{22}+\frac{3}{4}0_{22}$
paternal	$S_{32}T_{23}=\frac{1}{2}T_{33}+\frac{1}{2}0_{33}$	$S_{22}T_{23}=\frac{1}{2}T_{23}+\frac{1}{2}0_{23}$	$S_{32}T_{22}=0_{32}$	$S_{22}T_{22}=0_{22}$

involved in successive generations (i.e. a father-son step) the result is an 0 matrix, indicating independence between the two specified relatives. But, maternal grandfather-grandson are genotypically related, having a probability of $\frac{1}{2}$ of sharing an identical gene.

Another interesting point to be noted is that:

$$T_{32}T_{23}=\begin{bmatrix}1 & 0\\ \frac{1}{2} & \frac{1}{2}\\ 0 & 1\end{bmatrix}\begin{bmatrix}p & q & 0\\ 0 & p & q\end{bmatrix}=\begin{bmatrix}p & q & 0\\ \frac{1}{2}p & \frac{1}{2} & \frac{1}{2}q\\ 0 & p & q\end{bmatrix}=T_{33}=T.$$

That is, the relationship between *paternal* grandmother and granddaughter is the same as that between mother and daughter as though the intermediate father were non-existent. That, of course, should be the case because the father always transmits the same gene he received from the grandmother. This product of two transition matrices is also equivalent to a "factorization" of the T matrix for autosmal genes into two steps: the first step is to specify the probabilities that a gene is transmitted from parent to offspring, and the second, the probabilities of forming the three genotypes with that one gene given.

CORRELATION, GENOTYPIC AND PHENOTYPIC

9. The classical method of calculating the genotypic correlation between relatives is first to derive the joint distribution (correlation table) of the relatives in the general population. This can be obtained by multiplying the rows of our final transition matrix for the relatives by the appropriate "initial" frequencies of the given conditions. The correlation table thus obtained need not be given here. Suffice to say that the results are in complete agreement with those given by Fisher (1918) for autosomal genes and those by Hogben (1932) for sex-linked genes.

10. The correlation between relatives, however, need not be calculated from their joint distribution in the population. Before we show how it may be obtained directly from the components of the transition matrix of the stated relatives, we will first make an observation concerning the *additive* property of the correlation coefficient under some restrictive conditions. It may be shown that if there are N_1 pairs of values in one population with certain mean values, variances, and correlation coefficient r_1 ; and if there are N_2 pairs of values in another population with the same mean values and variances but correlation coefficient r_2 , then the correlation between the $N_1 + N_2$ pairs of values in the pooled population is

$$r = \frac{N_1}{N_1 + N_2} r_1 + \frac{N_2}{N_1 + N_2} r_2 .$$

That is, the correlation in the pooled population is the weighted mean of the separate correlations. Or, in other words, the pooled correlation may be subdivided into two components, corresponding to the contributions of the component populations. This analysis may obviously be extended to any number of component populations as long as they have the same mean and variance.

11. Now, consider the transition matrix for a specified type of relatives. Its general expression is, for autosomal genes,

$$c_I I + c_T T + c_0 0$$

where c_I , c_T , c_0 are the probabilities that the specified relatives should have both, one, or none genes identical by descent ($c_I + c_T + c_0 = 1$). The I-component contributes perfect (unity) correlation between the genotypes of the relatives while the 0-component contributes nothing to their correlation. If r_T is the correlation for the T-component (to be calculated after converting T into absolute frequencies), then the

total correlation between the relatives in the general population will simply be

$$r = c_I + c_T r_T$$

on account of the additive property of correlation coefficient described above. Therefore, to calculate the total r, we need only to calculate the one basic correlation r_T once and for all. Multiplying the three rows of T by p^2, $2pq$, q^2, it will be found that, as is well known, that $r_T = \frac{1}{2}$. Hence, for example, the correlations between

$$\text{grandparent-grandchild}, r = 0 + \frac{1}{2} r_T = \frac{1}{2}\left(\frac{1}{2}\right) = \frac{1}{4}.$$

$$\text{double first cousins}, \qquad r = \frac{1}{16} + \frac{6}{16} r_T = \frac{1}{16} + \frac{6}{16}\left(\frac{1}{2}\right) = \frac{1}{4}.$$

It should be noted that although these two types of relatives have the same correlation coefficient, the joint frequency distribution (correlation table) of these two types of relatives in the general population are quite different; for, in the former case, the transition matrix is $\frac{1}{2}T + \frac{1}{2}0$; in the latter, $\frac{1}{16}I + \frac{6}{16}T + \frac{9}{16}0$. This situation is exactly analogous to that of parent-offspring vs. full sibs.

12. One great advantage of subdividing the correlation between relatives into two components, as given in the preceding section, is that when A is dominant to a, the above formula for r still holds except that $r_T = \frac{1}{2}$ is to be replaced by $r_T^* = q/(1 + q)$, the latter being the correlation in the T-component population with dominance (Li, 1954, Chap. 2). For example, the phenotypic correlation between full sibs is

$$r^*_{\text{sibs}} = \frac{1}{4} + \frac{1}{2} r_T^* = \frac{1}{4} + \frac{1}{2}\left(\frac{q}{1 + q}\right)$$

in agreement with the more conventional expression $(1 + 3q)/4(1 + q)$. It assumes the value 5/12 when $q = \frac{1}{2}$, as obtained by Yule (1906, cf. Fisher, 1918). The phenotypic correlation between other types of relatives may be obtained in a similar manner.

13. For sex-linked genes the procedure of deriving the correlations is exactly the same as before except that now there are four basic correlations, corresponding to T_{33} , T_{32} , T_{32}, T_{22} components. Without dominance,

$$r_{T_{33}} = \tfrac{1}{2}, \qquad r_{T_{32}} = r_{T_{32}} = \sqrt{\tfrac{1}{2}}, \qquad r_{T_{22}} = 0.$$

For example, the transition matrix of full sisters being $S_{33} = \frac{1}{2}I_{33} + \frac{1}{2}T_{33}$, the correlation between full sisters with respect to sex-linked traits is

$$r_{\text{sisters}} = \tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2}) = \tfrac{3}{4}.$$

Another example is that the correlation for maternal uncle-niece, with transition matrix $\frac{1}{4}T_{23} + \frac{3}{4}0_{23}$ (Table 1), will be $\frac{1}{4}r_{T_{23}} = \frac{1}{4}\sqrt{\frac{1}{2}}$. The correlations arrived at this way are also in complete agreement with those obtained by classical method or by path coefficients.

14. For sex-linked traits with dominance, the phenotypic correlations for the four T-component populations are

$$r^*_{T_{33}} = \frac{q}{1+q}, \qquad r^*_{T_{32}} = r^*_{T_{23}} = \sqrt{\frac{q}{1+q}}, \qquad r^*_{T_{22}} = 0.$$

Thus, the phenotypic correlation between

$$\text{Full sisters: } r^* = \frac{1}{2} + \frac{1}{2}\left(\frac{q}{1+q}\right)$$

$$\text{Maternal uncle-niece: } r^* = \frac{1}{4}\sqrt{\frac{q}{1+q}}.$$

Note that for sex-linked genes, *whether with dominance or not*,

$$r_{T_{33}} = r^2_{T_{32}} = r^2_{T_{23}}.$$

Dominance, however, does not effect the correlation between two male relatives.

MULTIPLE ALLELES

15. The foregoing methods and results may be applied directly to multiple alleles, autosomal or sex-linked, without any change except that the matrices assume dimensions of 6×6, 6×3 or 3×3 for the case of three alleles with six genotypes. We shall give just one example to show the general validity of the method. In the following, the order of rows and columns are A_1A_1, A_1A_2, A_1A_3, A_2A_2, A_2A_3, A_3A_3 for autosomal genotypes or females if the genes are sex-linked. In the latter case, the order of males are A_1, A_2, A_3. The gene frequencies are p, q, r, respectively, where $p + q + r = 1$. The autosomal parent-offspring transition matrix is $T_{66} = T$. For sex-linked genes, the mother-son matrix is T_{63} and the father-daughter matrix is T_{36}. The paternal grandmother-granddaughter transition matrix is thus,

$$T_{63}T_{36} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p & q & r & 0 & 0 & 0 \\ 0 & p & 0 & q & r & 0 \\ 0 & 0 & p & 0 & q & r \end{bmatrix}$$

$$= \begin{bmatrix} p & q & r & 0 & 0 & 0 \\ \frac{1}{2}p & \frac{1}{2}p + \frac{1}{2}q & \frac{1}{2}r & \frac{1}{2}q & \frac{1}{2}r & 0 \\ \frac{1}{2}p & \frac{1}{2}q & \frac{1}{2}p + \frac{1}{2}r & 0 & \frac{1}{2}q & \frac{1}{2}r \\ 0 & p & 0 & q & r & 0 \\ 0 & \frac{1}{2}p & \frac{1}{2}p & \frac{1}{2}q & \frac{1}{2}q + \frac{1}{2}r & \frac{1}{2}r \\ 0 & 0 & p & 0 & q & r \end{bmatrix} = T_{66} = T.$$

We see that exactly the same relation as before holds. It may be also verified that $T_{66}^2 = \frac{1}{2}T_{66} + \frac{1}{2}0_{66}$ where every row of 0_{66} consists of p^2, $2pq$, $2pr$, q^2, $2qr$, r^2. In brief, all of the foregoing relations between the matrices hold for multiple alleles.

DISCUSSION

16. For autosomal genes, full sibs and double first cousins, being able to share both genes identical by descent, are called "bilineal" relatives (Cotterman, 1941), while all other types of relatives considered here are "unilineal" relatives, who can at most share one identical gene in common. For sex-linked genes, however, due to the haploid nature of the males, the I component appears not only in full sibs of the same sex (S_{33} and S_{22}) but also in two male relatives with a female as an intermediate connecting individual, such as maternal grandfather-grandson and maternal uncle-nephew relations (Table 1).

17. In formulating the relationship between two individuals, however, care must be taken to see that the formulation is valid biologically. To illustrate what we mean by this, let us consider the autosomal relationship given in Fig. 2. We see that the two individuals X and Y are half-sibs and therefore, when X is given, the conditional probabilities for Y is T^2. Now, Y and Z are also half-sibs and hence, for specified Y, the probabilities of the genotypes of Z are elements of T^2. But, when X is given, the probabilities for Z's genotype is *not* $T^2T^2 = T^4$, as might have been first supposed. Obviously, X and Z are independent individuals, for X is the child of A and B while Z is the child of C and

D, where the four parents involved are independent individuals. As a matter of fact, X is not even related with C, still less with a child of C by a second marriage. Therefore, in using the transition matrices, it should be borne in mind that one's relative's relatives are not necessarily still relatives. It seems that the rules for using T should be the same as that for using path coefficients. The latter method is beyond the scope of the present discussion; it has been given by Wright (e.g. 1934) and explained by Li (1954) in detail.

18. It should also be pointed out that the subdivision of the correlation coefficient between relatives as given by $r = c_I + c_T r_T$ is different from the subdivision facilitated by the method of path coefficients. The latter isolates the components of correlation due to each independent common ancestor of the two relatives concerned. Thus, the correlation between double first cousins, analyzed by path coefficients, will resolve into 4 *equal components* of magnitude 1/16 each, corresponding to the equal contributions of each of the four grandparents. In this connection, particular attention should be directed to the components of correlation between full sibs. Either method yields $r = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. According to our subdivision, the first $\frac{1}{4}$ is due to identical genotypes while the second $\frac{1}{4}$ is due to half-identical genotypes. By method of path coefficients, the first $\frac{1}{4}$ is due to the influence of one parent (mother, say) while the second $\frac{1}{4}$ is due to the influence of the other parent (father). In this special case it is incidental that the two systems of subdivision should yield the same apparent results.

19. One more application of the method of labeling and tracing the identity of independent genes may be mentioned. It concerns with the genotypic proportions of the offspring of consanguineous matings in a random mating population. The half-sibs of Fig. 2 may be employed as an example. Let the parents A, B, C be (1, 2), (3, 4), (5, 6), respectively. The child X may thus be (1, 3), (1, 4), (2, 3), (2, 4), while

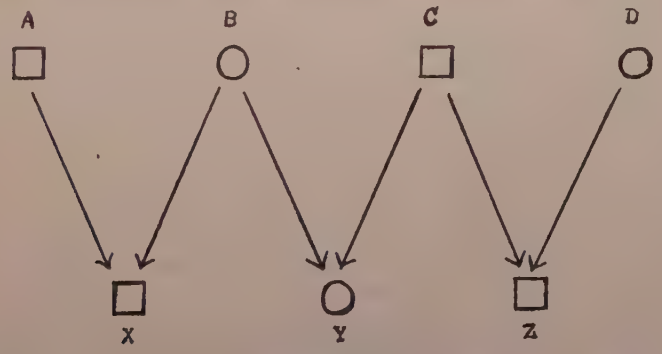


FIG. 2. THE INDEPENDENCE BETWEEN THE INDIVIDUAL X AND HIS HALF-SIB'S (Y's) HALF-SIB (Z)

his half-sib Y may be (3, 5), (3, 6), (4, 5), (4, 6). If the half-sibs X and Y mate, the sixteen possibilities of their offspring may be represented by the random union of 1, 2, 3, 4 from X with 3, 4, 5, 6 from Y. There will be $\frac{1}{16}$ (3, 3) and $\frac{1}{16}$ (4, 4) among the offspring. This 1/8 offspring is certainly homozygous because their two genes are derived from the same gene of their common grandparent. The probability of its being AA is p and being aa is q. The remaining 7/8 offspring is formed of independent genes, such as (1, 3), (1, 4), etc. Hence, the offspring of half-sib matings consists of

$$\begin{array}{ccc} AA & Aa & aa \\ \frac{7}{8}p^2 + \frac{1}{8}p & \frac{7}{8}(2pq) & \frac{7}{8}q^2 + \frac{1}{8}q \end{array}$$

in agreement with results obtained through other methods. The 1/8 in the above expression is usually designated by F, known as the "inbreeding coefficient" of the resultant offspring. It is the probability that the two genes received by an offspring be identical (derived from the same gene of a common ancestor of his parents).

ACKNOWLEDGEMENTS

The authors are indebted to Dr. Antonio Ciocco in suggesting the problem of employing stochastic matrix to derive the frequencies of and the correlations between relatives. They also wish to thank Dr. Bentley Glass for his helpful consultation in interpreting the results and Dr. Donovan Thompson for his reading this paper in manuscript.

REFERENCES

1. Cotterman, C. W. Relatives and human genetic analysis. *Sci. Monthly 53*: 227–234. 1941.
2. Crow, J. F. Breeding structure of populations. *Statistics and Mathematics in Biology*, 543–556, Iowa State College Press, 1954.
3. Feller, William. *An introduction to probability theory and its applications*. Chap. 5 and 15. Wiley, N.Y. 1950.
4. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin. 52*: 399–433. 1918.
5. Hogben, L. Filial and fraternal correlations in sex-linked inheritance. *Proc. Roy. Soc. Edin. 52*: 331–336. 1932.
6. Hogben, L. A matrix notation for Mendelian populations. *Proc. Roy. Soc. Edin. 53*: 7–25. 1933.
7. Li, C. C. *Population Genetics*. Chap. 2, 3, 4, and 13. University of Chicago Press (in press). 1954.
8. Malecot, G. *Les Mathematiques de L'Heredite*. Masson & Cie, Paris. 1948.
9. Wright. S. The biometric relations between parent and offspring. *Genet. 6*: 111–123. 1921.
10. Wright, S. The method of path coefficients. *Ann. Math. Stat. 5*: 161–215. 1934.

A STOCHASTIC MODEL FOR THE SELECTION OF MACRONUCLEAR UNITS IN *PARAMECIUM* GROWTH*

A. W. Kimball and A. S. Householder
Oak Ridge National Laboratory

1. *Introduction*

Experimental evidence obtained by various research workers [see Sonneborn, (1947) and (1949)] has provided interesting information about the existence and behavior of macronuclear units in *Paramecium.* All paramecia of the same stock probably have about the same number of macronuclear units. Before division takes place each macronuclear unit doubles, and each daughter cell presumably selects one-half of the macronuclear units present at the time of division. In this way the number of units per animal is kept constant in a manner analogous to the behavior of chromosomes. Unlike chromosomes, however, the macronuclear units do not separate into homologous sets and as yet very little is known about the nature of the macronuclear unit selection process. The purpose of this paper is to test one hypothesis about the selection process by incorporating it into a stochastic model and by comparing predicted results with experimental data.

In section 2, a state is defined and the transition probabilities are derived. Section 3 deals with the theory which underlies the algebraic treatment of finite Markov chains. Numerical results for relatively small numbers of macronuclear units are given in section 4, and these results are compared with experimental data in section 5.

2. *The transition probabilities*

The state of an animal is defined to be the type distribution of its macronuclear units (hereafter referred to simply as units). If an animal has n units each unit may be different from every other unit, or all units may be alike, and any combination between these two extremes is also possible. The rth state, defined as a particular distribution of unit types, may be represented conveniently by the vector,

$$(1) \qquad a_r = (a_{r1}, a_{r2}, \cdots, a_{rn}) \qquad r = 1, \cdots, m,$$

where m is the number of different states and where each a_{ri} $(i = 1, \cdots, n)$ is the number of units of one type. Clearly each a_{ri}

*Presented at the American Statistical Association Annual Meeting in Washington, 1953 in a session held jointly with The Biometric Society (ENAR) and the Institute of Mathematical Statistics.

may assume any integral value from 0 to n subject to the condition that

$$\sum_{i=1}^{m} a_{ri} = n.$$

In (1), the order of types is disregarded. For example if $n = 3$, the state (2, 1, 0) implies that the macronucleus contains two units of one type, one of a second type and none of the third type. No distinction is made between two macronuclei, one having two A-units, one B-unit and no C-units and the other having two B-units, one A-unit and no C-units. Two animals having such macronuclei are said to be in the same state, (2, 1, 0).

Readers familiar with number theory will recognize that the number of possible states, m, is identically equal to the number of partitions of n. Thus the number of possible states is the coefficient of t^n in the power series expansion of

$$\prod_{p=1}^{\infty} (1 - t^p)^{-1}.$$

No workable formula for m as a function of n has ever been derived although much work has been done on asymptotic formula and recurrence relations.

For certain experimental problems information is needed as to the change in the distribution of states in a *Paramecium* population as the population grows. If a population is generated from a single animal whose state is known and if the probability of transition from one state to any other state as a result of division can be computed, the distribution of states in the population after ν divisions can be obtained by treating the growth process as a finite Markov chain. Likewise if a population is generated from a group of animals and if the initial distribution of states is known, the distribution after ν divisions may be derived directly from the results obtained for a population generated from a single animal. The development throughout the paper assumes that all animals in the population undergo the same number of divisions per unit length of time. While this is most certainly not true for any specific time period, experimental evidence indicates that the assumption becomes more closely approximated as the period of observation increases. In principle division rates could be incorporated into the problem by approximating the distribution from experimental data and by using this information in establishing the transition probabilities. Whether this approach would be feasible is questionable.

If an animal normally has n units in its macronucleus, then at the time of division $2n$ units are available to the daughter cells each of

which ultimately receives n of them. One possible mechanism for the manner in which a daughter selects n out of the $2n$ units is the hypothesis of completely random selection. According to this hypothesis each of the $\binom{2n}{n}$ ways of selecting n units out of $2n$ units is assumed to be equally probable. For the purpose of computing transition probabilities only one daughter need be considered, since the second daughter's lot is uniquely determined after the first daughter's selection is made. In general, the probability of transition from state a_r to state a_s is given by

$$p_{rs} = \frac{1}{\binom{2n}{n}} \left[\frac{(2a_{r1})!\,(2a_{r2})! \cdots (2a_{rn})!}{a_{s1}!\,a_{s2}! \cdots a_{sn}!\,\alpha!\,\beta! \cdots \sigma!} \right] \dagger x_{ij} \dagger \tag{2}$$

$$(i, j = 1, \cdots, n; r, s = 1, \cdots, m),$$

where $\alpha, \beta, \cdots, \sigma$ are the numbers of like a_{sj} and

$$x_{ij} = \frac{1}{(2a_{ri} - a_{sj})!}. \tag{3}$$

The symbol $\dagger x_{ij} \dagger$ refers to the permanent, a mathematical form identical to the determinant except that all terms are taken with the positive sign. In (3), if $a_{sj} > 2a_{ri}$, x_{ij} is taken to be zero, consistent with the usual definition of the factorial of a negative integer.

The validity of (2) can be proved rigorously but it can also be deduced almost directly from a simple example. Consider the case $n = 3$. Here $m = 3$ and the possible states are (1, 1, 1), (2, 1, 0) and (3, 0, 0). In general we are interested in the probability of transition from state $a_r = (a_{r1}, a_{r2}, a_{r3})$ to state $a_s = (a_{s1}, a_{s2}, a_{s3})$. If the elements of a_s are all different, the number of ways state a_s may be obtained from state a_r is

$$\begin{aligned} &\binom{2a_{r1}}{a_{s1}}\binom{2a_{r2}}{a_{s2}}\binom{2a_{r3}}{a_{s3}} + \binom{2a_{r1}}{a_{s1}}\binom{2a_{r2}}{a_{s3}}\binom{2a_{r3}}{a_{s2}} + \binom{2a_{r1}}{a_{s2}}\binom{2a_{r2}}{a_{s1}}\binom{2a_{r3}}{a_{s3}} \\ &+ \binom{2a_{r1}}{a_{s2}}\binom{2a_{r2}}{a_{s3}}\binom{2a_{r3}}{a_{s1}} + \binom{2a_{r1}}{a_{s3}}\binom{2a_{r2}}{a_{s1}}\binom{2a_{r3}}{a_{s2}} + \binom{2a_{r1}}{a_{s3}}\binom{2a_{r2}}{a_{s2}}\binom{2a_{r3}}{a_{s1}}, \end{aligned} \tag{4}$$

which may be written in the form,

$$\frac{(2a_{r1})!(2a_{r2})!(2a_{r3})!}{a_{s1}!a_{s2}!a_{s3}!} \left[\begin{aligned} &x_{11}x_{22}x_{33} + x_{11}x_{23}x_{32} + x_{12}x_{21}x_{33} \\ &+ x_{12}x_{23}x_{31} + x_{13}x_{21}x_{32} + x_{13}x_{22}x_{31} \end{aligned} \right], \tag{4'}$$

where the x_{ij} are defined as in (3). The expression in brackets in (4')

is immediately identified as the permanent $\dagger x_{ij} \dagger$, since it contains exactly the six terms of the corresponding third order determinant but with all signs positive. The expression in (4′) was derived on the assumption that the a_{sj} were all different. If in fact some of them are alike, we must make the usual correction which consists of dividing by the factorials of the numbers of like a_{sj}. Finally after division by $\binom{2n}{n}$, we arrive at the desired probability given in (2). The extension to cases in which $n > 3$ is almost intuitively obvious.

The statement that the transition probabilities can be computed by considering only one daughter may not at first glance appear valid. For example if $n = 4$, there are five possible states, (1, 1, 1, 1), (2, 1, 1, 0), (2, 2, 0, 0), (3, 1, 0, 0) and (4, 0, 0, 0). An animal in state (3,1, 0, 0) can upon division give rise to daughters which are in states (2, 2, 0, 0), (3, 1, 0, 0) or (4, 0, 0, 0). If from the six units of type A and 2 units of type B which are available one daughter selects 2 of type A and 2 of type B, the remaining daughter will be left with 4 of type A and none of type B. That is, the two daughters will be in different states, (2, 2, 0, 0) and (4, 0, 0, 0). This might appear to make our calculations invalid. Fortunately, however, each result of this kind is exactly balanced by another selection which occurs with exactly the same frequency. In this case if the first daughter selects 4 of type A, the second daughter is left with two of type A and two of type B, so that the two daughters are again in the states (2, 2, 0, 0) and (4, 0, 0, 0). The net result is that the relative frequency of occurrence of each state is exactly what it would be if daughter pairs were always alike. For many transitions, of course, the two daughters from a single animal actually are in the same state.

3. *The general theory*

Since reproduction does not increase the number of distinct types of units, a daughter cell can have at most as many types as the parent, and there is a finite probability that the number will be smaller. The total number of possible states is m, the number of partitions of n, as indicated already. These states can be divided into n groups, according to the number of distinct types of units present. Let the groups be designated $g_1, \cdots, g_n$, where g_i is the group of those possible states for which exactly i distinct types are present. Then no descendant of a cell of group g_i can be in a group g_j for $j > i$.

The vector $a_r = (a_{r1}, a_{r2}, \cdots, a_{rn})$, where the a_{ri} are non-negative integers satisfying

$$\sum_i a_{ri} = n,$$

is taken to represent the state in which there are a_{r1} elements of one type, a_{r2} of another, etc. It is no restriction to suppose that

$$a_{r1} \geq a_{r2} \geq \cdots \geq a_{rn} .$$

We can make a lexicographic ordering of the possible states and let the subscripts r correspond to this ordering, as follows. If $r < s$, then for some i,

$$0 = a_{rn} - a_{sn} = a_{r,n-1} - a_{s,n-1} = \cdots = a_{r,i+1} - a_{s,i+1} > a_{ri} - a_{si} .$$

Thus in state a_1 , all elements are alike; in state a_2 , all but one are alike; in state a_3 , there are two elements of one type and $n - 2$ of another; $\cdots$; in state a_n , the elements are all different. We are interested in determining, as a function of ν, the probability that a daughter cell of the νth generation is in group g_1 , or, more generally, in g_i or some lower group.

Let p represent the matrix of probabilities p_{rs} . Let $p_{0;r}$ represent the probability that some particular cell is in state a_r , and let p_0 represent the row vector whose elements are these probabilities. Then

$$p_1 = p_0 p$$

is a row vector whose sth element is the probability that a particular daughter cell is in state a_s . More generally,

$$p_\nu = p_0 p^\nu$$

is a row vector whose sth element is the probability that a particular daughter cell of the νth generation is in state a_s . We recall that all elements of p and all elements of each p_ν are non-negative, and that

$$\sum_s p_{\nu;s} = \sum_s p_{rs} = 1.$$

The classification into groups of states provides a natural partitioning of the matrix p and, in fact, it has the form,

$$p = \begin{pmatrix} q_1 & 0 & 0 & \cdots \\ q_{21} & q_2 & 0 & \cdots \\ q_{31} & q_{32} & q_3 & \cdots \\ \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & \end{pmatrix} ,$$

where the submatrices above the diagonal are all null because of the impossibility of a transition to a higher group. The principal submatrix q_1 is a scalar and, in fact, $q_1 = 1$ since group g_1 consists of the single state in which all elements are alike. The order of q_2 is either $n/2$ or

else $(n - 1)/2$, whichever is an integer. Both q_n and q_{n-1} are scalars, q_{n-2} is of second order, and q_{n-3} is of third order*.

If, for any square matrix A, we let $A(\lambda)$ represent the characteristic function, then in this notation,

$$p(\lambda) = q_1(\lambda)\ q_2(\lambda) \cdots q_n(\lambda).$$

Since $q_1(\lambda) = 1 - \lambda$, it follows that $\lambda = 1$ is a characteristic root of p. In a matrix of positive elements, the greatest characteristic root cannot exceed the greatest sum of the elements of any one row and cannot be exceeded by the least of these. But every row of p beyond the first has at least one non-null element that is not an element of any q_i . Hence, $q_i(1) \neq 0$ for every $i > 1$, and 1 is therefore a simple root of $p(\lambda) = 0$ and exceeds in magnitude every other root. Thus if $\lambda_1 = 1$, λ_2 , λ_3 , $\cdots$, λ_μ are the distinct characteristic roots of p,

$$|\lambda_t| < 1,\ t = 2, \cdots, \mu.$$

Let Λ represent the Jordan normal form of p, so that

$$p = M^{-1}\Lambda M$$

for some non-singular M. In Λ each diagonal element is a λ_t , the first subdiagonal consists of ones and zeros, and all other elements are null. Then

$$p^\nu = M^{-1}\Lambda^\nu M$$

and

$$p_\nu = p_0 M^{-1}\Lambda^\nu M.$$

Each element of the vector p_ν can be expressed as a linear combination of the λ_t^ν $(t = 1, \cdots, \mu)$ whose coefficients are either constants or are polynomials in ν. If Λ is a diagonal matrix the coefficients are constant. If λ_t is a root of multiplicity m_t , then the coefficient of λ_t^ν is a polynomial in ν of degree at most m_t . Hence

$$p_\nu = k_1 + k_2(\nu)\lambda_2^\nu + \cdots + k_\mu(\nu)\lambda_\mu^\nu ,$$

where the $k_t(\nu)$ are vectors whose elements are either constants or polynomials in ν.

For increasing ν each row of p^ν approaches a characteristic row vector associated with the largest characteristic root of p, and each column approaches a characteristic column vector associated with this same characteristic root. However the largest characteristic root is

*For $n < 6$, q_{n-2} and q_{n-3} may be of lower order.

$\lambda_1 = 1$, and the associated characteristic row vector is the unit vector

$$\rho_1 = (1, 0, \cdots, 0).$$

Since $p_\nu \to k_1$ as $\nu \to \infty$ and since the sum of the elements is always unity, it follows that $k_1 = \rho_1$, so that

$$p_\nu = \rho_1 + k_2(\nu)\lambda_2^\nu + \cdots + k_\mu(\nu)\lambda_\mu^\nu .$$

We are interested in the rate at which $p_\nu \to \rho_1$. If there is a root λ_2 which exceeds in modulus all the other roots except λ_1 then asymptotically,

$$p_\nu \doteq \rho_1 + k_2(\nu)\lambda_2^\nu .$$

For the special cases $n \leq 7$ it turns out that $\lambda_2 = (2n - 2)/(2n - 1)$, and this is a simple root. Hence for these cases, $k_2(\nu)$ is a constant independent of ν. This root satisfies $q_2(\lambda) = 0$. For $n \leq 6$ it is also true that the next largest root is $\lambda_3 = (2n - 4)/(2n - 1)$, which is again a simple root and satisfies $q_3(\lambda) = 0$. The matrix q_2 may be expected to have the largest root different from unity since its row sums will be greatest. It remains to be determined whether these formulas are valid generally.

Although it has not been determined that $\lambda_2 = (2n - 2)/(2n - 1)$ in general, nevertheless we can set a bound for this root. The matrix q_{21} is, in fact, a column vector whose jth element can be written

$$\frac{\binom{2n - 2j}{n}}{\binom{2n}{n}} = \frac{n(n - 1) \cdots (n - 2j + 1)}{2n(2n - 1) \cdots (2n - 2j + 1)}.$$

The smallest of these elements is the last for which $j = [\frac{n}{2}]$, and is equal to $1/\binom{2n}{n}$ if n is even and $(n + 1)/\binom{2n}{n}$ if n is odd. If we call this γ_n, then $1 - \gamma_n$ is the greatest row sum for the matrix q_2 and $1 - (n - 1)/(4n - 2) = (3n - 1)/(4n - 2)$ is the least. Hence λ_2 is bounded between these two limits.*

For more complete information one would like, of course, to have general formulas for all the λ's and the corresponding k's. Intuitively the nature of the model would seem to rule out the possibility for oscillatory behavior, and we might therefore expect that all λ_i would be positive and real. This is borne out in all cases examined ($n \leq 6$), the roots actually being positive rationals. A few of the roots have been determined in general: q_n and q_{n-1} are both scalars having the values

*Dr. Motoo Kimura, one of the referees, has shown that $\lambda_2 = (2n - 2)/(2n - 1)$ holds asymptotically.

$$q_n = \frac{2^n}{\binom{2n}{n}}, \qquad q_{n-1} = \frac{2^{n-1}\binom{n+1}{1}}{\binom{2n}{n}};$$

q_{n-2} is of order 2 and has the roots

$$\frac{2^n}{\binom{2n}{n}}, \quad \frac{2^{n-2}\binom{n+2}{2}}{\binom{2n}{n}};$$

and q_{n-3} is a matrix of order 3 and has the roots

$$\frac{2^n}{\binom{2n}{n}}, \quad \frac{2^{n-1}\binom{n+1}{1}}{\binom{2n}{n}}, \quad \frac{2^{n-3}\binom{n+3}{3}}{\binom{2n}{n}}.$$

The roots of q_i , as i decreases by unit steps from $i = n$, form a pattern which suggests that the roots of q_{n-4} would be

$$\frac{2^n}{\binom{2n}{n}}, \quad \frac{2^{n-2}\binom{n+2}{2}}{\binom{2n}{n}}, \quad \frac{2^{n-4}\binom{n+4}{4}}{\binom{2n}{n}}, \quad \frac{2^{n-6}\binom{n+6}{6}}{\binom{2n}{n}},$$

and similarly for the remaining q_i . Although we can offer no proof of this, the formulas do provide the correct roots for $n \leq 6$.

4. *Numerical results for some special cases*

To illustrate how the computations are carried out, we shall consider the case $n = 5$ in detail. In this case $m = 7$, and the possible states are

$$a_1 = (5, 0, 0, 0, 0)$$
$$a_2 = (4, 1, 0, 0, 0)$$
$$a_3 = (3, 2, 0, 0, 0)$$
$$a_4 = (3, 1, 1, 0, 0)$$
$$a_5 = (2, 2, 1, 0, 0)$$
$$a_6 = (2, 1, 1, 1, 0)$$
$$a_7 = (1, 1, 1, 1, 1).$$

The first step is the calculation of p, the matrix of transition probabilities. From (2), for example, we find

$$\binom{10}{5} p_{77} = \frac{2!2!2!2!2!}{1!1!1!1!1!5!} \begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{vmatrix}^{+} = 32,$$

and

$$\binom{10}{5} p_{64} = \frac{4!2!2!2!0!}{3!1!1!0!0!2!2!1!} \begin{vmatrix} 1 & 1/6 & 1/6 & 1/24 & 1/24 \\ 0 & 1 & 1 & 1/2 & 1/2 \\ 0 & 1 & 1 & 1/2 & 1/2 \\ 0 & 1 & 1 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 & 1 \end{vmatrix}^{+} = 48.$$

In evaluating the permanents, two properties are useful. Constants may be factored from rows or columns as in determinants, and the permanent of an $n \times n$ array consisting of all ones is $n!$. Repeated application of (2) to obtain all the transition probabilities yields

$$p = \frac{1}{252} \begin{bmatrix} 252 & 0 & 0 & 0 & 0 & 0 & 0 \\ 56 & 140 & 56 & 0 & 0 & 0 & 0 \\ 6 & 66 & 180 & 0 & 0 & 0 & 0 \\ 6 & 60 & 40 & 80 & 66 & 0 & 0 \\ 0 & 12 & 56 & 64 & 120 & 0 & 0 \\ 0 & 6 & 12 & 48 & 90 & 96 & 0 \\ 0 & 0 & 0 & 0 & 60 & 160 & 32 \end{bmatrix}.$$

The next step is the determination of the characteristic roots and vectors of p. Three roots may be obtained directly and the other four are solutions of the two equations,

$$\begin{vmatrix} (80/252 - \lambda) & 66/252 \\ 64/252 & (120/252 - \lambda) \end{vmatrix} = 0,$$

$$\begin{vmatrix} (140/252 - \lambda) & 56/252 \\ 66/252 & (180/252 - \lambda) \end{vmatrix} = 0.$$

Hence, the roots* are

$$(\lambda_1 , \lambda_2 , \cdots , \lambda_7) = 1/252(252, 96, 224, 32, 168, 96, 32).$$

The characteristic vectors, which may be obtained in numerous ways, are

$$M = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \\ m_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -3 & 2 & 0 & 0 & 0 & 0 \\ -25 & 11 & 14 & 0 & 0 & 0 & 0 \\ 1 & -5 & 0 & 10 & 0 & -10 & 4 \\ 19 & -27 & -30 & 16 & 22 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 \\ -1 & 5 & -10 & 0 & 30 & -40 & 16 \end{bmatrix},$$

where m_i is the characteristic vector corresponding to λ_i .

At this point, given any initial probability distribution of states p_0 , we could compute the distribution after ν generations,

$$p_\nu = p_0 M^{-1} \Lambda^\nu M,$$

after first computing the inverse of M. (In this case Λ is a diagonal matrix the elements of which are the λ_i .) For the purpose at hand (see section 5), we are interested only in the special case in which all cells initially are in the state a_7 , i.e., the state in which all units are different. This corresponds to the initial distribution $p_0 = (0, 0, 0, 0, 0, 0, 1)$. Thus it is clear that we need only the last row of the inverse of M which is

$$p_0 M^{-1} = (1, -35/64, 5/64, 9/68, 15/204, 5/2, 1/34).$$

Furthermore (see section 5), we are interested only in the first element of p_ν which is the probability corresponding to the state a_1 . In other words this element gives the probability that, after ν generations, a daughter cell which originally came from a cell in which all units were different would have all its units alike. The element is

$$1 - (35/64)(96/252)^\nu - (125/64)(224/252)^\nu + (9/68)(32/252)^\nu + (285/204)(168/252)^\nu - (1/34)(32/252)^\nu.$$

Expressions giving this probability have been derived for $n = 2$,

*Ordered to correspond with the rows of p.

3, 4, 5 and are shown in Table 1. The corresponding survival curves have been computed and are shown in Figure 1. The dotted lines in Figure 1 represent experimental data which are discussed in the next section.

TABLE 1.

The probability that, after ν generations, a daughter cell will have all units alike, given that the original parent cell had all units different.

Units per Cell (n)	Probability
2	$1 - (2/3)^{\nu}$
3	$1 - (3/2)(4/5)^{\nu} + (1/2)(2/5)^{\nu}$
4	$1 - (39/22)(30/35)^{\nu} + (20/35)^{\nu} - (5/22)(8/35)^{\nu}$
5	$1 - (125/64)(56/63)^{\nu} + (95/68)(42/63)^{\nu}$ $- (35/64)(24/63)^{\nu} + (7/68)(8/63)^{\nu}$

5. *A comparison with experimental data*

The data which provide a test of the proposed model for macronuclear unit selection actually came from an experiment with *Tillina* carried out* by Dr. Josephine Bridgman. It seems certain that the macronucleus of *Tillina* is of basically the same structure as that of its near relative *Paramecium*. Thus we are justified in using the data from this organism for our comparison.

In the experiments considered here, tillinas were irradiated at a dose which resulted in essentially 100 per cent mortality after about twelve divisions. Such severe early mortality suggests that each of the n units was rendered lethal as a result of irradiation. If this is true and if we assume that all animals started out in the state $(1, 1, \cdots, 1)$, the survival curves which would be expected for various numbers of units are precisely those shown in Figure 1. That is, these curves represent the expected proportion of a population which would not be in state $(n, 0, \cdots, 0)$ after ν divisions. If each unit is lethal the state $(n, 0 \cdots, 0)$ would be synonomous with death.

The results of two experiments plotted as per cent survival against number of divisions are shown as the dotted lines in Figure 1. If the model and assumptions are correct, the number of macronuclear units would have to be in the neighborhood of 3 or 4. Since the actual number of units is likely to be 20 or greater, the model or the additional as-

*Work performed in the Biology Division, Oak Ridge National Laboratory.

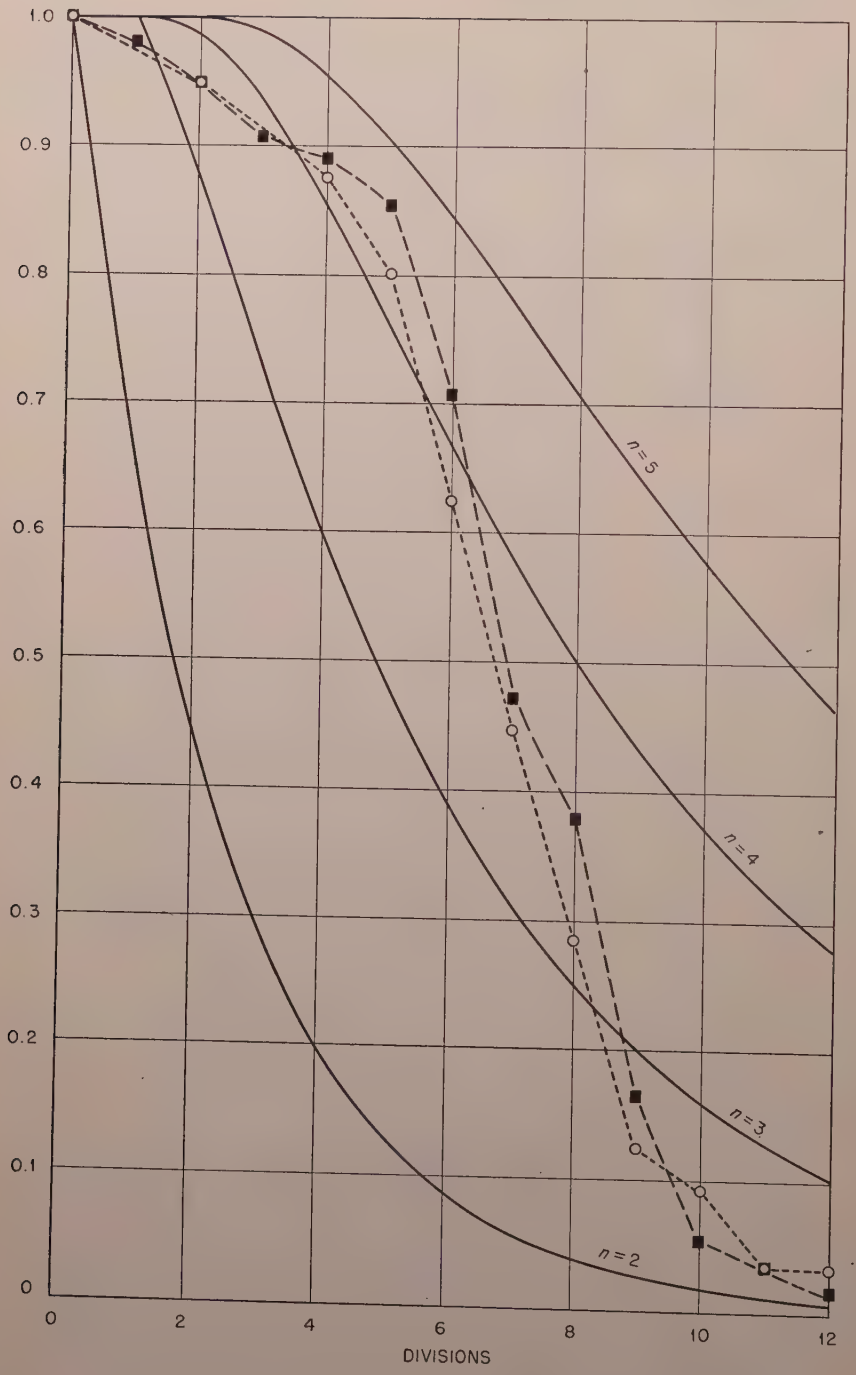


Fig. 1. Probability of Survival.

sumptions, or both, must be discarded. Clearly any change in the assumption concerning the initial distribution of states would result in a higher estimate of n. That is, if any of the original animals already had some of their units alike, the predicted survival curves would be uniformly lower since in each case the state $(n, 0, \cdots, 0)$ would be reached more rapidly by the animals starting out with some units alike. This would be one way to obtain a better fit to the model. Actually, however, among less than one hundred units, it is not very likely that any two would mutate lethally in exactly the same way. For this reason the assumption that all animals are initially in the state $(1, 1, \cdots, 1)$ is probably a close approximation to reality.

One reasonable assumption which would bring the theory and experiment closer together is that some number of like units less than n would produce death. Previous experience with other organisms has suggested that with recessive lethal mutations and a high degree of polyploidy, death can result from states of less than complete homozygosis. Thus, if we assume that death can result if all but k of the n units are alike, k may be varied until the theoretical curve for, say, $n = 20$ approaches the experimental curve. If the value of k required to bring the two curves together is too large, say greater than 50 per cent of n, this approach would also have to be discarded since it is not likely that death would result if only half of the units were alike. These computations will not be practicable until more complete results are obtained for the general case.

Finally, one could consider various changes in the basic model. Allowance might be made for spontaneous mutations, or some hypothesis other than complete random selection might be proposed. Unfortunately, in this experiment some evidence is available which suggests that the macronuclear unit selection hypothesis may not be the proper way to describe the lethal effects of radiation. In some respects the death phenomenon seems to be an all-or-none affair in the sense that death for an animal in any particular line of descent can be predicted rather well after the first division has occurred. Furthermore, division rates are not uniform and seem to depend on whether the animal is destined to expire. If these observations are correct, the proposed model may never fit the data very well, and even if it does, its validity would be questionable. Nevertheless, the stochastic process itself is rather general in nature and perhaps has counterparts in other biological problems or even in other fields of endeavor.

6. *Summary and conclusions*

A stochastic model is proposed for the selection of macronuclear units in *Paramecium* growth. The model presupposes random selection and that death can result only in the completely homozygotic state. Although full results for the general case have not been obtained, enough has been done to indicate that the model must be changed if reasonable agreement with experimental data is to be achieved. The most plausible alteration which would provide better agreement is an allowance for death in states somewhat less than complete homozygosis. Before computations to test this hypothesis can be started, more results are needed for the general case of n macronuclear units. At present complete results are available only for values of n up to 6. Hand computation for larger values of n rapidly becomes impracticable.

The authors are indebted to Dr. R. F. Kimball, Biology Division, Oak Ridge National Laboratory, for suggesting this problem and assisting in its formulation and for a careful review of the manuscript. We also wish to thank Dr. Bridgman for permission to use her hitherto unpublished data.

REFERENCES

1. Sonneborn, T. M. "Recent advances in the genetics of *Paramecium* and *Euplotes*", *Advances in Genetics*, 1: 263–358. 1947.
2. Sonneborn, T. M. "Ciliated protozoa: cytogenetics, genetics and evolution", *Ann. Rev. Microbiology*, 3: 55–80. 1949.

INCOMPLETE BLOCK RANK ANALYSIS: ON THE APPROPRIATENESS OF THE MODEL FOR A METHOD OF PAIRED COMPARISONS.[1]

RALPH ALLAN BRADLEY
Virginia Agricultural Experiment Station of the
Virginia Polytechnic Institute

1. *Introduction*:

The application of any statistical technique to numerical data is strictly appropriate only when the data conform to the assumptions and requirements inherent in the development of that statistical method. In the analysis of variance, we require normality, additivity, independence, and homogeneity of variances of the data. When these conditions are not at least approximately satisfied, the analysis is suspect. In postulating the mathematical model for the rank analysis of incomplete block designs and more specifically for the method of paired comparisons [1] we chose the model because it seemed intuitively plausible, because it was related to the binomial model, and because it was mathematically workable.

The major objective of this paper is to develop a procedure for testing the appropriateness of the model for the method of paired comparisons. The test is then applied to a variety of experiments involving taste, preference, and appearance judgments for a technique is only very useful when it is applicable to a considerable variety of data.

In the reference cited above, three tests of significance were proposed. Two of them are tests of treatment effects, and the other is a test of agreement among judges or over groups of the data. The test of agreement could be regarded as a test of the need of the model that permits judges to differ on their judgments as opposed to the model in which it is postulated that all judges are coordinated in their judgments on the attributes under consideration. We could call the test of agreement a between-group test of goodness of fit.

These comments will become clearer in view of the following summary of the mathematical model for paired comparisons.

---

[1]Prepared under a Research and Marketing Act Contract, Project RM: c–629–1, with the Bureau of Agricultural Economics.

2. *The model and tests of hypotheses*:

Let us consider the method of paired comparisons. A basic repetition of a paired comparisons experiment may be defined to be a set of incomplete blocks of size two in which every pair of t items are compared once and only once. The experiment is taken to consist of n such basic repetitions. Various procedures for the analysis of paired comparisons are available. The method of analysis used depends on the form in which data are recorded: measurements or scores may be available for each sample or for the difference between items in each pair, or only the preferred item in each pair may be indicated, thus constituting a ranking procedure. Other innovations, such as the grouping of repetitions of the experiment by judges or by time or by some other relevant characteristic, may be developed for further experimental control. We shall review the formulation of the mathematical model and the test procedures for a method of paired comparisons, the rank analysis of incomplete block designs, wherein items are only ranked and some control grouping of repetitions of the experiment may be present.

In summarizing the model for the method of paired comparisons under consideration, we shall first assume that the experiment consists of n repetitions for t treatments and that the repetitions are a homogeneous set. Then it is postulated that associated with each of the t treatments, say $T_1, \cdots, T_t$, there exist parameters, π_i for T_i, such that $\pi_i \geqq 0$ and $\sum_{i=1}^{t} \pi_i = 1$. The latter restriction is one that is imposed to insure determinancy. The behavior of the parameters is further defined with the probability statement that, if X_i generally denotes an observation on a sample of T_i,

$$P(X_i > X_j) = \pi_i/(\pi_i + \pi_j) \tag{1}$$

in the comparison of T_i with T_j.

Experimental observation is limited to ranking in the pairs. Thus, r_{ijk} is defined to be the rank of T_i in the block in which T_i is compared with T_j in the kth repetition of the experiment, $1 \leqq k \leqq n$. A non-null decision is required in each pair and of necessity r_{ijk} is either one or two with $r_{ijk} + r_{jik} = 3$. The rank of one will normally be assigned to that treatment of a pair that is judged 'superior' on the basis of the test attribute of the treatments.

The element of the likelihood function for the block containing T_i and T_j may be written as

$$\pi_i^{2-r_{ijk}} \pi_j^{2-r_{jik}} (\pi_i + \pi_j)^{-1}.$$

The general likelihood function for the t treatments and n homogeneous repetitions is abbreviated through use of the functional form $L(\pi_i)$ with

$$(2) \qquad L(\pi_i) = \prod_i \pi_i^{2n(t-1)-\sum_k \sum_i' r_{ijk}} \prod_{i<j} (\pi_i + \pi_j)^{-n}$$

where the index of summation or of multiplication covers the full range when not otherwise specified, where the prime indicates that the index of summation j takes all values except i, and where the second product contains one factor representing each treatment pair. The maximum likelihood estimator of π_i is denoted by p_i. Procedures for obtaining these estimators are outlined in two papers [1, 4]. We shall need to refer to (2) in developing the theory for testing the appropriateness of the model.

In some cases the model that assumes homogeneous repetitions is not realistic. This is particularly true in certain types of taste-testing experiments where a number of judges perform repetitions of the experiment and where the main objective is the detection of treatment differences. Then it may be reasonable to postulate the existence of parameters $\pi_{1u}, \cdots, \pi_{tu}, \pi_{iu} \geqq 0, \sum_i \pi_{iu} = 1$ for the uth of g groups of repetitions in the experiment. This simply means that the experiment may be designed to contain g groups of repetitions with n_u repetitions in the uth group, $\sum_u n_u = n$. Repetitions within a group are taken to be homogeneous, but group differences in treatment parameters are permitted. The mathematical model differs from the simpler case described only in the grouping of the repetitions. The likelihood function in this case is a simple product of g functions of the form (2). The maximum likelihood estimator of π_{iu} may be designated by p_{iu}.

The test procedures that have been developed under these models for paired comparisons are as follows:

$$\text{Test (i):} \quad H_0 : \pi_i = 1/t \qquad \text{for all } i,$$
$$H_1 : \pi_i \neq 1/t \qquad \text{for some } i.$$

The likelihood ratio test depends on the statistic[2]

$$(3) \quad B_1 = n \sum_{i<j} \log (p_i + p_j) - \sum_i \left\{2n(t-1) - \sum_k \sum_i' r_{ijk}\right\} \log p_i .$$

When we let λ_1 be the likelihood ratio statistic for test (i),

$$(4) \qquad -2 \ln \lambda_1 = nt(t-1) \ln 2 - 2B_1 \ln 10$$

has for large values of n the χ^2-distribution with $(t-1)$ degrees of

[2] We shall use log and ln to represent logarithms to base 10 and base e respectively.

freedom[3]. This first test is a test of treatment equality under the assumption of homogeneous repetitions.

$$\text{Test (ii):} \qquad H_0 : \pi_{iu} = 1/t \qquad \text{for all } i \text{ and } u,$$
$$H_2 : \pi_{iu} \neq 1/t \qquad \text{for some } i \text{ and } u.$$

We define B_1^u to correspond to B_1 of (3) for the n_u homogeneous repetitions of the uth group and

$$B_1^c = \sum_u B_1^u . \tag{5}$$

The likelihood ratio statistic λ_1^c for Test (ii) is defined by

$$-2 \ln \lambda_1^c = nt(t-1) \ln 2 - 2B_1^c \ln 10 \tag{6}$$

which has for large values of $n_1 , \cdots , n_g$ the χ^2-distribution with $g(t-1)$ degrees of freedom[4].

$$\text{Test (iii):} \qquad H_0 : \pi_{iu} = \pi_i \qquad \text{for all } i \text{ and } u,$$
$$H_3 : \pi_{iu} \neq \pi_i \qquad \text{for some } i \text{ and } u.$$

Test (iii) may be regarded as a test of agreement of treatment parameters over the g groups, and this is analogous to a test of treatment $\times$ group interaction. In this case, we may define the likelihood ratio statistic λ_a by

$$\begin{aligned} -2 \ln \lambda_a &= -2(\ln \lambda_1^c - \ln \lambda_1) \\ &= 2(B_1 - B_1^c) \ln 10 \end{aligned} \tag{7}$$

where λ_1 is obtained by assuming that all g groups of repetitions may be pooled to form an experiment with n homogeneous repetitions. The large-sample distribution of $-2 \ln \lambda_a$ is that of χ^2 with $(g-1)(t-1)$ degrees of freedom, and the distribution for small samples is dependent on $\pi_1 , \cdots , \pi_t$ as nuisance parameters.

3. *Tests of goodness of fit*:

We shall limit our main consideration to a test of goodness of fit for the simple paired comparisons model for t treatments and n homogeneous repetitions of the possible pairwise treatment comparisons.

---

[3]The exact distribution of B_1 has been tabled for small values of t and n. Tables for $t = 3$, $n = 1, \cdots , 10$ and for $t = 4$, $n = 1, \cdots , 6$ have been published [1]. Unpublished tables for $t = 4$, $n = 7, 8$, and for $t = 5$, $n = 1, \cdots , 5$ have been prepared.

[4]Limited tables for the distribution of B_1^c are published [1]. Only cases where $n_1 , \cdots , n_g$ are all equal are considered.

It will then be easy to extend this theory to the more general model with grouped repetitions.

The most general model available for a paired comparisons experiment within a group of data is one in which each comparison is treated independently as a small binomial-type experiment. One parameter is then estimated for each comparison. For example, when treatments T_i and T_j are compared, a parameter $\bar{\pi}_{ij}$, the probability that T_i is ranked above T_j , would be estimated. The complementary probability $\bar{\pi}_{ji} = 1 - \bar{\pi}_{ij}$. Now, if t treatments are involved in an experiment of this type, $\binom{t}{2}$ parameters must be estimated. Estimates are obtained from relative frequencies so that

$$\bar{p}_{ij} = f_{ij}/n \tag{8}$$

where $\bar{p}_{ij}$ is the estimator of $\bar{\pi}_{ij}$, f_{ij} is the number of times that T_i ranks above T_j , and n is the number of times such a comparison is made. This general model fits the experimental data exactly in the sense that observed and expected frequencies of a particular ranking are identical.[5] When $\binom{t}{2}$ parameters are to be estimated, the likelihood function may be written as

$$\bar{L}(\bar{\pi}_{ij}) = \prod_{ij} \bar{\pi}_{ij}^{f_{ij}} \bar{\pi}_{ji}^{f_{ji}} \tag{10}$$

where

$$f_{ij} + f_{ji} = n \qquad \text{and} \qquad \bar{\pi}_{ij} + \bar{\pi}_{ji} = 1.$$

To test the appropriateness or the goodness of fit of the model for paired comparisons for homogeneous repetitions, we are interested in the following test:

Test of Fit I: $H_0 : \bar{\pi}_{ij} = \pi_i/(\pi_i + \pi_j), \quad i \neq j, i, j = 1, \cdots, t$

$H_I : \bar{\pi}_{ij} \neq \pi_i/(\pi_i + \pi_j)$ for some i and j.

An approximate test is possible using the likelihood ratio procedure.

The likelihood ratio statistic for Test of Fit I depends on $\bar{L}(\bar{p}_{ij} \mid H_0)$ and $\bar{L}(\bar{p}_{ij} \mid H_I)$ where $\bar{L}$ is defined by (10) and these functions represent evaluations of $\bar{L}$ in terms of estimators $\bar{p}_{ij}$ obtained under the assumptions of H_0 and H_I respectively. Further, it will be noted that $\bar{L}(\bar{p}_{ij} \mid H_0) \equiv L(p_i)$ where L is defined in (2). These quantities are

---

[5] The more restrictive model of Section 2 is obtained when we set $\bar{\pi}_{ij} = \pi_i/(\pi_i + \pi_j)$. Further, the relative frequencies are related to the sums of ranks previously defined for

$$\sum_k \sum_j{}' r_{ijk} = 2n(t-1) - \sum_j{}' f_{ij} \,. \tag{9}$$

When it can be done without ambiguity, we shall use Σr_i as an abbreviation for the left-hand member of (9) and refer to it as the 'total sum of ranks for T_i'.

evaluated using the appropriate maximum likelihood estimators and we have

$$\ln \overline{L}(\bar{p}_{ij} \mid H_0) = -B_1 \ln 10, \tag{11}$$

and

$$\ln \overline{L}(\bar{p}_{ij} \mid H_I) = \sum_{i \neq j} f_{ij} \ln f_{ij} - n\binom{t}{2} \ln n \tag{12}$$

in view of (3) and (8). B_1 is the statistic tabled for Test (i) and is available without special calculation for a fairly wide range of values of t and of n. For the test of goodness of fit, the statistic,

$$-2 \ln \lambda_I = 2\left[\sum_{i \neq j} f_{ij} \ln f_{ij} - n\binom{t}{2} \ln n + B_1 \ln 10 \right], \tag{13}$$

has the χ^2-distribution with $\binom{t}{2} - t + 1$ degrees of freedom for large values of n.

The computation for the test is most easily carried out after forming the two-way table commonly used in tabulating the results for methods of paired comparisons. For three items or treatments, this table appears as follows.

TABLE I. TABLE FOR THE TEST OF GOODNESS OF FIT

	T_1	T_2	T_3	Σr_i
T_1	—	f_{12}	f_{13}	$\Sigma r_1 = 4n - f_{12} - f_{13}$
T_2	f_{21}	—	f_{23}	$\Sigma r_2 = 4n - f_{21} - f_{23}$
T_3	f_{31}	f_{32}	—	$\Sigma r_3 = 4n - f_{31} - f_{32}$

When the repetitions of the paired comparisons experiment are grouped into g groups, the uth of which contains n_u repetitions, $u = 1, \cdots, g$, $\sum_u n_u = n$, separate and possibly distinct sets of parameters, $\bar{\pi}_{iju}$ and π_{iu} are assumed to exist within each group. Statistics f_{iju} and $\sum r_{iu}$ are computed for each of the g groups. Within-group tests of goodness of fit are made as described for Test I and the test statistics may be designated by $-2 \ln \lambda_{Iu}$. We note that all of these parameters and statistics correspond to those of the simpler case and that we have only added the additional subscript u to designate the groups of the experiment.

The test specification for an over-all test of goodness of fit is

Test of Fit II:

$$H_0 : \bar{\pi}_{iju} = \pi_{iu}/(\pi_{iu} + \pi_{ju}) \quad i \neq j, i, j = 1, \cdots, t, u = 1, \cdots, g.$$

$$H_{II} : \bar{\pi}_{iju} \neq \pi_{iu}/(\pi_{iu} + \pi_{ju}) \quad \text{for some } i \text{ and } j \text{ and } u.$$

In view of Test of Fit I and the fact that rankings by repetitions of the experiment are taken to be independent in probability, the likelihood ratio statistic for Test of Fit II is

$$-2 \ln \lambda_{II} = -2 \sum_{u=1}^{g} \ln \lambda_{Iu}, \tag{14}$$

and this statistic has the χ^2-distribution with $g\{\binom{t}{2} - t + 1\}$ degrees of freedom for large values of n_u.

Test of Fit II should be used when repetitions of the experiment are grouped. When grouping is employed, it must have been necessary to assume a priori that one set of parameters, $\pi_1, \cdots, \pi_t$, would not be sufficient to characterize the experimental results. It would not thus be reasonable to use Test of Fit I and to ignore the grouping in these situations.

4. *The tests of goodness of fit associated with expected frequencies*:

In tests of goodness of fit, the procedures usually involve computing expected cell frequencies, say f'_{ij}, and forming χ^2 by taking sums of terms of the form, $(f_{ij} - f'_{ij})^2/f'_{ij}$. Again, in this case, we can show that our tests, which depend on logarithms of cell frequencies, are approximately equivalent to the more usual ones. We shall consider Test of Fit I, and the results obtained will apply directly to the second test as well.

Sums of ranks are directly related to observed cell frequencies through (9) and Table I. Expected cell frequencies are related to the estimators $p_1, \cdots, p_t$ through

$$f'_{ij} = np_i/(p_i + p_j). \tag{15}$$

Then, with the use of (3) and (13), $B_1 = -\sum_{i \neq j} f_{ij} \log (f'_{ij}/n)$, and

$$-2 \ln \lambda_I = 2 \sum_{i \neq j} f_{ij} \ln (f_{ij}/f'_{ij}). \tag{16}$$

It will be convenient for the moment to write $f_{ij}/f'_{ij} = 1 + \epsilon_{ij}$, where ϵ_{ij} may be either positive or negative. Now,

$$-2 \ln \lambda_I = 2 \sum_{i \neq j} f'_{ij}(1 + \epsilon_{ij}) \ln (1 + \epsilon_{ij}).$$

We use power series expansions for the logarithms stopping with the second terms. The errors committed in so doing will not be large

if $| \epsilon_{ij} |$ is small. Difficulty is introduced, however, if $\epsilon_{ij} = -1$ which occurs when an observed cell frequency is zero. Upon expanding the logarithms and noting that $\sum_{i \neq j} f'_{ij} \epsilon_{ij} = 0$, we have

$$-2 \ln \lambda_I \approx \sum_{i \neq j} f'_{ij} \epsilon^2_{ij} - \sum_{i \neq j} f'_{ij} \epsilon^3_{ij} .$$

The final result is that

$$-2 \ln \lambda_I \approx \sum_{i \neq j} \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}} \tag{17}$$

if we neglect the terms involving ϵ^3_{ij} and note the relationship of ϵ_{ij} with cell frequencies.

The procedure for the second test of goodness of fit is clear since the test statistic is made up of a sum of terms like the left-hand member of (17). We shall retain again the definition of the symbols used in (17) but add another subscript u to specify the group of repetitions. Then.

$$-2 \ln \lambda_{II} \approx \sum_{u=1}^{g} \sum_{i \neq j} \frac{(f_{iju} - f'_{iju})^2}{f'_{iju}}. \tag{18}$$

The two methods of computing χ^2 for tests of goodness of fit or of the appropriateness of the models will usually be equivalent for practical purposes. Further, there is little difference in computing time required when values of B_1 are available. Otherwise, the second forms, (17) and (18), may be slightly easier to handle. Both methods require values of $p_1 , \cdots , p_t$, which must be obtained from tables or computed for the experiment. The additional work in making these tests of the basic models over the analysis of the data for test purposes is very small.

We have noted that difficulty is encountered in obtaining the forms (17) and (18) if observed cell frequencies are small. In addition, the χ^2-tests developed are only approximate and require reasonably large numbers of repetitions. If one were about to appraise the applicability of the model of the method of paired comparisons to a specific type of experimental problem on which a considerable amount of experimentation is planned, it would be very useful to conduct a fairly extensive experiment to obtain data for tests of goodness of fit.

5. *Applications of the tests of goodness of fit to experimental data*:

In the following subsections, we indicate the results of testing the appropriateness of the model postulated for the rank analysis of incomplete block designs for three different types of data. These sets of data are those that happen to be available in the form required for the test. No deliberate selection has been made, and the reader may

assess the results as they were obtained. The data were obtained by the credited research workers exactly as shown, and we have only obscured the quantitative nature of the treatments because the data have not yet been published elsewhere.

It does seem necessary that some treatment differences exist before data can show any marked departures from the model used. Thus it is to show that these are not simply 'blank' experiments that the levels of significance for tests of treatment effects have been included under each of the following tables. These significance levels should not be confused with significance levels associated with values of χ^2 also given under the tables.

(i) Experiment 1. The Effects of Peanuts in Hogs' Rations on the Flavor of Fresh Pork Roasts. C. M. Kincaid, H. R. Thomas, and Lyle L. Davis, Virginia Agricultural Experiment Station.

We shall present the data from this experiment summarized in tables similar to Table I. The treatments consist of different rations fed to hogs that were selected for similar characteristics and from a limited number of litters. A is a corn ration, B is a corn ration supplemented by a small peanut ration, and C is a corn ration supplemented by a larger peanut ration than was the case for B. The experiment was conducted so that the tasters recorded a rank of 1 for that sample in a pair that they thought came from the animal fed the larger amount of peanuts. Members of the panel were obtained after preliminary training and selection, and it was believed that they would be able to recognize the effect of a peanut supplemented diet on fresh roast pork. Actually these data do not reveal any over-all treatment effects but are somewhat indicative of detectable differences among roasts in the sets.

During the conduct of the experiment, even after the preliminary training, it appeared that there was still some difficulty in coordinating the rankings of the judges. For this reason, the combined analysis [Test (ii)] was used for tests of treatment differences and we now test the model for each judge separately. It also appeared that differences among roasts might be as important as differences between rations, and hence we also test the model for each set of roasts over the panel of judges. Five judges and five sets of roasts were used in the experiment.

The required tables for these ten tests of the model are shown below. The calculations required are illustrated in detail for the first judge, and we note that each χ^2 computed in Experiment 1 has 1 degree of freedom. In each case here χ^2 has been calculated using both of the forms (13) and (17) with the latter value being shown in parentheses. The 5 per cent significance level of χ^2 with 1 degree of freedom is 3.84 and the 1 per cent level is 6.63.

TABLE II. RESULTS FOR JUDGE I

	A	B	C	Σr_i
A	–	7 (6.00)	5 (6.00)	28
B	3 (4.00)	–	6 (5.00)	31
C	5 (4.00)	4 (5.00)	–	31

$t = 3$, $n = 10$, $B_1 = 8.8560$, $\chi_1^2 = 1.24$, (1.23)
Sig. level of treat. = .78.

Using Test of Fit I in the form (13), we have

$$-2 \ln \lambda_I = \chi_1^2 = 2(2.3026)[7 \log 7 + 5 \log 5 + 3 \log 3 + 6 \log 6 + 5 \log 5 + 4 \log 4 - 30 \log 10 + 8.8560] = 1.24.$$

This value is shown below Table II followed by the value obtained using expected cell frequencies.

To use the second form (17), we require $p_A = .4286$, $p_B = .2857$, and $p_C = .2857$ obtained from tables [1]. The expected cell frequencies are shown in parentheses in Table II and they were calculated using the relationship, (15). The alternative method of computing χ^2 yields

$$\chi_1^2 = \frac{(7-6)^2}{6} + \frac{(5-6)^2}{6} + \frac{(3-4)^2}{4} + \frac{(6-5)^2}{5} + \frac{(5-4)^2}{4} + \frac{(4-5)^2}{5} = 1.23.$$

TABLE III.
RESULTS FOR JUDGE II

	A	B	C	Σr_i
A	–	3	4	33
B	7	–	5	28
C	6	5	–	29

$t = 3$, $n = 10$, $B_1 = 8.6191$,
$\chi_1^2 = 0.15$ (0.16)
Sig. level of treat. = .50.

TABLE IV.
RESULTS FOR JUDGE III

	A	B	C	Σr_i
A	–	5	3	32
B	5	–	5	30
C	7	5	–	28

$t = 3$, $n = 10$, $B_1 = 8.7973$,
$\chi_1^2 = 0.57$ (0.57)
Sig. level of treat. = .63.

In all cases considered, the two methods of computing χ^2 for Test of Fit I yielded values in close agreement.

TABLE V.
RESULTS FOR JUDGE IV

	A	B	C	Σr_i
A	—	3	4	33
B	7	—	5	28
C	6	5	—	29

$t = 3, n = 10, B_1 = 8.6191,$
$\chi_1^2 = 0.15\ (0.16)$
Sig. level of treat. = .50.

TABLE VI.
RESULTS FOR JUDGE V

	A	B	C	Σr_i
A	—	4	4	32
B	6	—	8	26
C	6	2	—	32

$t = 3, n = 10, B_1 = 8.3162,$
$\chi_1^2 = 1.37\ (1.35)$
Sig. level of treat. = .25.

Test of Fit II may be applied here to the over-all experiment. Then, using (14), we have

$$-2 \ln \lambda_{II} = \chi_5^2 = 1.24 + 0.15 + 0.57 + 0.15 + 1.37 = 3.48$$

TABLE VII.
RESULTS FOR ROASTS A_1, B_1, C_1

	A	B	C	Σr_i
A	—	5	3	32
B	5	—	2	33
C	7	8	—	25

$t = 3, n = 10, B_1 = 7.8787$
$\chi_1^2 = 0.19\ (0.19)$
Sig. level of treat. = .10.

TABLE VIII.
RESULTS FOR ROASTS A_2, B_2, C_2

	A	B	C	Σr_i
A	—	7	6	27
B	3	—	8	29
C	4	2	—	34

$t = 3, n = 10, B_1 = 8.2548$
$\chi_1^2 = 2.33\ (2.31)$
Sig. level of treat. = .21.

TABLE IX.
RESULTS FOR ROASTS A_3, B_3, C_3

	A	B	C	Σr_i
A	—	1	5	34
B	9	—	8	23
C	5	2	—	33

$t = 3, n = 10, B_1 = 6.6647$
$\chi_1^2 = 0.32\ (0.32)$
Sig. level of treat. = .007.

TABLE X.
RESULTS FOR ROASTS A_4, B_4, C_4

	A	B	C	Σr_i
A	—	6	3	31
B	4	—	4	32
C	7	6	—	27

$t = 3, n = 10, B_1 = 8.6191$
$\chi_1^2 = 0.55\ (0.56)$
Sig. level of treat. = .50.

The level of significance is approximately .6 indicating good agreement of the model and the observations.

When the experiment is regrouped by sets of roasts, A_1 , B_1 , C_1 ; $\cdots$; A_5 , B_5 , C_5 , the following tables and analyses are obtained.

TABLE XI.

RESULTS FOR ROASTS A_5 , B_5 , C_5

	A	B	C	Σr_i
A	—	3	3	34
B	7	—	7	26
C	7	3	—	30

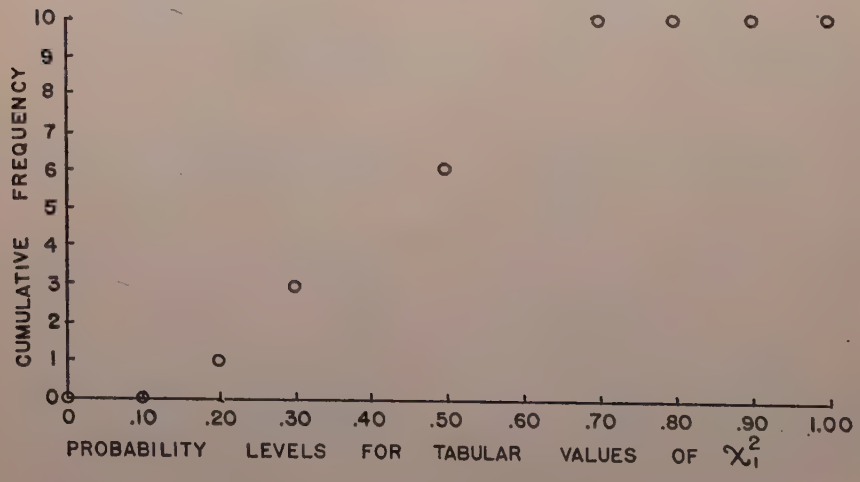
$t = 3, n = 10, B_1 = 8.0688\ \chi_1^2 = 0.51\ (0.51)$
Sig. level of treat. = .13.

For this grouping, $-2 \ln \lambda_{II} = 3.90$ and, with 5 degrees of freedom, this value of χ^2 is at the .57 level of significance.

Figure I shows the observed cumulative frequencies of values of χ_1^2 plotted against the probability of a larger χ_1^2 as obtained using tables

FIGURE 1

CUMULATIVE FREQUENCIES OF OBSERVED VALUES OF χ_1^2 THAT EXCEED TABULAR VALUES OF χ_1^2 AT SPECIFIED SIGNIFICANCE LEVELS



of χ^2. The values of χ_1^2 computed are clearly in line with what would be expected for a sample of ten observations on a χ_1^2-variate. These results are indicative of the appropriateness of the model for this experiment. We note that there is some dependence among the ten computed values of χ_1^2 that are used in Figure I.

(ii) Experiment 2. The Effect of Color Set on Preference, Red Color, and Yellow Color for Stayman Apples. G. E. Mattus, Virginia Agricultural Experiment Station.

A commercial hormone preparation called 'color set' is used to delay the dropping of mature apples from the trees. The preparation is applied as a spray a short time before harvest to prolong the picking period. It was thought that the color set produced some change in the surface color of the apples and in particular that it increased the intensity of the yellow tones and also created the appearance of more red color. The following experiment was conducted to investigate these possibilities.

Two groups of judges were asked to rank representative plates of Stayman apples for preference, red color, and yellow color in a paired comparisons experiment. The treatments were A (check or no color set), B (q p.p.m. of color set), and C (m p.p.m. of color set) with $q > m$. Seven judges, who were selected from the stenographic and laboratory staff, were used in one group and eight horticulturists were used in a second group[6]. The rankings of the judges in each group were pooled making totals of 21 and 24 repetitions of the paired comparison experiment.

The results of tests of goodness of fit are summarized in two-way tables as they were for Experiment 1. Since the values of n are 21 and 24, values of p_A , p_B , and p_C were obtained by direct calculation

TABLE XII.
COLOR PREFERENCE—GP. 1

	A	B	C	Σr_i
A	—	12	6	66
B	9	—	3	72
C	15	18	—	51

$t = 3$, $n = 21$, $B_1 = 15.5184$,
$\chi_1^2 = 0.43$.
Sig. level of treat. $< .005$.

TABLE XIII.
COLOR PREFERENCE—GP. 2

	A	B	C	Σr_i
A	—	17	8	71
B	7	—	8	81
C	16	16	—	64

$t = 3$, $n = 24$, $B_1 = 19.8594$,
$\chi_1^2 = 1.38$.
Sig. level of treat. $= .016$.

[6]The results of a third group composed of statisticians have been deleted as unreliable for they showed extreme disagreement on color preferences.

(cf [4]). Here again the values of χ^2 are not suggestive of poor agreement of the model with the data. One value, that in Table XVII, is rather large but this may be partly due to small cell frequencies.

TABLE XIV.
MOST RED COLOR—GP. 1

	A	B	C	Σr_i
A	—	10	8	66
B	11	—	13	60
C	13	8	—	63

$t = 3, n = 21, B_1 = 18.7155,$
$\chi_1^2 = 1.30.$
Sig. level of treat. = .57.

TABLE XV.
MOST RED COLOR—GP. 2

	A	B	C	Σr_i
A	—	13	7	76
B	11	—	12	73
C	17	12	—	67

$t = 3, n = 24, B_1 = 21.1635,$
$\chi_1^2 = 2.11.$
Sig. level of treat. = .31.

TABLE XVI.
MOST YELLOW COLOR—GP. 1

	A	B	C	Σr_i
A	—	11	5	68
B	10	—	5	69
C	16	16	—	52

$t = 3, n = 21, B_1 = 16.3254,$
$\chi_1^2 = 0.01.$
Sig. level of treat. < .005.

TABLE XVII.
MOST YELLOW COLOR—GP. 2

	A	B	C	Σr_i
A	—	14	0	82
B	10	—	4	82
C	24	20	—	52

$t = 3, n = 24, B_1 = 13.2043,$
$\chi_1^2 = 6.58.$
Sig. level of treat. < .005.

(iii) Experiment 3. The Effect of Monosodium Glutamate (M.S.G.) on the Flavor of Applesauce and Dehydrated Apple Slices. Lyle L. Davis, Virginia Agricultural Experiment Station.

A fairly extensive experiment was conducted on the effect of various amounts of monosodium glutamate in applesauce and dehydrated sliced apples. Monosodium glutamate is frequently used as a taste stimulant. Much of the experiment was conducted by the method of paired comparisons and evaluated by means of the method of the rank analysis of incomplete block designs. The score cards were immediately summarized by computing treatment sums of ranks and unfortunately the greater part of these cards were then destroyed.

The following tables show results for only a part of the experiment. The treatments may be taken to be A (u p.p.m. of m.s.g.), B (v p.p.m.

of m.s.g.), C (w p.p.m. of m.s.g.) and D (check or no m.s.g.). For all sections of data presented $u < v < w$ but for the four sections u, v, and w are not identical concentrations.

In each group each judge performed just one repetition of the paired comparisons design. The results of all judges in each group were pooled for the analysis and the tables below show the cell frequencies for the groups together with values of B_1 and χ^2, the latter for tests of goodness of fit. χ^2 now has three degrees of freedom. Judges recorded that sample of a pair that they preferred in each case.

TABLE XVIII.
M.S.G. IN APPLESAUCE—GP. 1

	A	B	C	D	Σr_i
A	—	3	3	3	15
B	1	—	3	4	16
C	1	1	—	0	22
D	1	0	4	—	19

$t = 4$, $n = 4$, $B_1 = 5.4139$, $\chi^2_3 = 6.94$.
Sig. level of treat. = .059.

TABLE XIX.
M.S.G. IN APPLESAUCE—GP. 2

	A	B	C	D	Σr_i
A	—	2	5	1	28
B	4	—	5	2	25
C	1	1	—	1	33
D	5	4	5	—	22

$t = 4$, $n = 6$, $B_1 = 8.1788$, $\chi^2_3 = 0.76$
Sig. level of treat. = .010.

TABLE XX.
M.S.G. IN APPLESAUCE—GP. 3

	A	B	C	D	Σr_i
A	—	2	2	3	17
B	2	—	2	2	18
C	2	2	—	2	18
D	1	2	2	—	19

$t = 4$, $n = 4$, $B_1 = 7.1154$, $\chi^2_3 = 0.54$.
Sig. level of treat. = .986.

TABLE XXI.
M.S.G. IN DEHYDRATED APPLE SLICES

	A	B	C	D	Σr_i
A	—	4	4	0	34
B	3	—	5	2	32
C	3	2	—	3	34
D	7	5	4	—	26

$t = 4$, $n = 7$, $B_1 = 11.2387$, $\chi^2_3 = 6.32$.
Sig. level of treat. = .104.

Here two values of χ^2_3 are approaching the 5% significance level (7.82) and two are quite small. Again, the large values appear where zero cell frequencies occur and in such cases χ^2 is likely to be a poorer approximation to the distribution of the appropriate likelihood ratio statistic than when small cell frequencies do not occur. Further difficulty may be introduced due to the small values of n. The mean value of χ^2_3 is expected to be 3 and we have an average value of 3.5 for this experiment.

6. *Concluding remarks:*

In using a test for the appropriateness of a model as we have done, it is customary to use a continuity correction in computing χ^2. This tends to diminish the computed value of χ^2 and hence to reduce its apparent significance. We have not used any such corrections in the values computed and they may be biased in such a way as to emphasize apparent departures from the model. However, in our data, twenty values of χ^2 have been computed and only one is significant at the 5 per cent level of significance.

It may be argued that the data presented are too scanty to demonstrate the applicability of the model for paired comparisons in general. The major objective of this paper is to present procedures for testing the models and to illustrate the applications of the procedures. As a secondary objective we have presented additional experimental results with a view to showing enough data to provide some assurance that the model may be appropriate for experiments involving subjective judgments. It has recently come to our attention that J. W. Hopkins has independently supervised extensive experiments to provide data for testing the appropriateness of the models for experiments on basic taste sensations. The results of his work have been reported [3] and he did not find any indication of serious departures of observations from the observed frequencies expected under the model.

The author would like to express his appreciation to Lyle L. Davis, G. E. Mattus, C. M. Kincaid and H. R. Thomas for the use of their experimantal results.

7. *References:*

1. R. A. Bradley and M. E. Terry. The Rank Analysis of Incomplete Block Designs. *I* The Method of Paired Comparisons. *Biometrika 39*, 324–345, 1952.

2. R. A. Bradley. Some Statistical Methods in Taste Testing and Quality Evaluation. *Biometrics 9*, 22–38, 1953.

3. J. W. Hopkins. Incomplete Block Rank Analysis: Some Taste Test Results. Unpublished Manuscript, 1954.

4. M. E. Terry, R. A. Bradley, and Lyle L. Davis. New Designs and Techniques for Organoleptic Testing. *Food Technology 6*, 250–254, 1952.

INCOMPLETE BLOCK RANK ANALYSIS: SOME TASTE TEST RESULTS[1]

J. W. Hopkins
Division of Applied Biology, National Research Council of Canada

INTRODUCTION

Bradley, Terry and co-workers (2, 3, 9) have put forward a solution of the problem of incomplete block rank analysis, and have applied it to the design and analysis of sensory difference tests of flavor and palatability. In its simplest form of paired comparisons in balanced incomplete blocks of 2, their method rests upon two basic assumptions. The first of these is that all rankings are statistically independent. The second is that if aliquots of m test samples are appraised by each of n subjects, there exist n sets of non-negative real numbers π_{1j} , π_{2j} , $\cdots$. π_{mj} $(j = 1, 2, \cdots n)$ termed "ratings" and each satisfying the defining condition $\sum_i \pi_{ij} = 1$, such that for all pairs of samples h and i, the probability of the jth subject ranking i above h in a single paired comparison is specified by $\pi_{ij}/(\pi_{hj} + \pi_{ij})$.

Appropriateness of this conceptual model in any specific instance is a question of fact. Technological interpretation of subjective test results also involves another factual consideration, namely the extent of inherent differences in the π_{ij} characteristic of individual subjects j. Three series of experimental results bearing upon these points are reported below.

EXPERIMENTATION

Eleven subjects were recruited from laboratory staff. *A*, *B*, *C*, *G* and *H* were women; *D*, *E*, *F*, *I*, *J* and *K* men. All had previously participated in taste tests.

In Expt. I, 6 subjects, *A*, *B*, *C*, *D*, *E* and *F* each made 20 independent replicate rankings of relative sweetness of the two test aliquots in each of 12 pairs of aqueous sugar solutions. Six of these comprised the possible pairings of solutions *S*1, *S*2, *S*3 and *S*4 containing respectively 2.16, 2.32, 2.49 and 2.67% sucrose by weight in tap water; the other 6 comprised the possible pairings of solutions *L*1, *L*2, *L*3 and *L*4 having 6.90, 7.25, 7.62 and 8.00% lactose by weight. These concentrations increase geometrically by factors of 1.073 for *S* and 1.051 for *L*. Preliminary trials indicated that they provided suitable degrees of taste contrast.

---

[1]N.R.C. Publication No. 3357.

TABLE I.

Individuals' Recorded Frequency f of Specified Rankings in 20 Appraisals, and Corresponding Bradley-Terry Expectations f'

Specified ranking	A		B		C		D		E		F		Total f
	f	f'	f	f'	f	f'	f	f'	f	f'	f	f'	
I.S. 2 > 1	16	15.8	17	16.3	15	15.6	12	13.8	16	14.3	19	16.6	94
3 > 1	17	17.9	18	17.8	19	18.8	19	18.0	13	16.2	18	19.5	105
4 > 1	20	18.8	19	19.8	20	19.8	20	19.3	18	18.5	20	19.9	118
3 > 2	13	13.9	14	13.0	15	16.4	16	16.1	16	12.6	19	17.9	92
4 > 2	17	16.2	19	19.3	20	19.2	17	18.6	17	16.6	19	19.4	109
4 > 3	11	13.0	20	18.8	16	17.0	16	15.2	15	14.9	15	15.7	93
L. 2 > 1	12	11.0	14	15.0	15	15.7	14	14.4	15	14.8	16	16.6	86
3 > 1	9	10.6	17	16.6	19	17.6	13	13.7	16	16.8	17	17.3	92
4 > 1	16	15.4	19	18.4	18	18.6	17	16.0	18	17.3	19	18.9	107
3 > 2	10	9.6	12	12.3	11	13.3	8	9.1	12	13.0	11	11.2	64
4 > 2	15	14.6	15	15.7	17	15.6	13	12.1	15	13.8	15	15.5	90
4 > 3	14	15.0	14	13.9	12	12.8	11	12.9	9	10.9	15	14.6	75

Specified ranking	B		E		G		H		I		J		Total f
	f	f'	f	f'	f	f'	f	f'	f	f'	f	f'	
II.S. 2 > 1	11	12.7	15	15.8	9	11.9	9	10.3	13	11.8	12	12.1	69
3 > 1	18	16.8	17	16.9	17	15.4	16	14.6	15	15.5	17	15.2	100
4 > 1	20	19.7	20	19.4	20	18.7	16	16.1	17	17.6	15	16.8	108
3 > 2	14	15.0	12	12.0	12	13.8	14	14.4	16	14.1	14	13.5	82
4 > 2	19	19.6	17	17.8	17	18.1	15	15.9	16	16.7	15	15.6	99
4 > 3	19	18.7	17	16.9	16	16.2	13	12.0	15	13.7	15	12.6	95
SM. 2 > 1	13	15.7	10	11.4	9	11.2	14	15.0	18	16.8	13	12.2	77
3 > 1	19	16.9	17	14.2	12	10.5	19	18.0	19	18.6	12	13.2	98
4 > 1	20	19.6	16	16.4	12	11.2	19	18.9	18	19.5	17	16.6	102
3 > 2	10	12.0	11	12.9	8	9.2	13	15.0	15	14.3	12	11.0	69
4 > 2	18	18.5	15	15.5	9	10.0	18	17.0	18	17.6	15	15.2	93
4 > 3	18	17.9	14	13.1	11	10.8	12	13.1	16	14.9	14	14.3	85

Specified ranking	B		D		E		G		H		K		Total
	f	f'	f	f'	f	f'	f	f'	f	f'	f	f'	
III. $W > X$	4	5.4	15	13.2	5	5.0	8	9.5	8	8.2	10	10.3	50
$W > Y$	16	14.2	17	18.2	3	3.1	9	6.5	7	5.5	5	6.0	57
$W > Z$	8	8.3	12	12.6	9	8.8	7	8.0	9	10.3	10	8.7	55
$X > Y$	17	17.3	17	16.8	8	7.1	8	7.0	6	7.1	5	5.8	61
$X > Z$	12	13.1	11	9.4	13	14.0	6	8.5	13	12.1	9	8.5	64
$Y > Z$	6	4.5	2	2.9	17	16.2	15	11.5	15	14.7	11	12.8	66

In Expt. II, subjects *B*, *E*, *G*, *H*, *I* and *J* also each made 20 independent replicate rankings of relative sweetness of the two test aliquots in each of 12 pairs. Six of these were identical with the foregoing pairs of *S*1, *S*2, *S*3 and *S*4; the other 6 were pairings of solutions *SM*1, *SM*2, *SM*3 and *SM*4 derived from them by addition to each of 0.003% quinine sulfate, 0.15% sodium chloride and 0.21% tartaric

acid to provide uniform accompanying bitter-salt-sour taste stimuli.

For Expt. III, four qualitatively different test solutions, *W*, *X*, *Y* and *Z* were used. These were arrived at from a basic mixture of 0.625% sucrose, 0.075% sodium chloride, 0.02675% tartaric acid and 0.00015% quinine sulfate by weight, again in tap water, by reducing the concentration of each of these constituents in turn by 1/4. Subjects *B*, *D*, *E*, *G*, *H* and *K* then made 20 independent preference rankings of these solutions in all 6 pairings.

In all experiments, each subject ranked six pairs of aliquots at daily mid-morning and mid-afternoon tasting sessions. The order of presentation of test pairs and of aliquots within pairs was objectively randomized subject to 2 restrictions. (i) Each pair of aliquots was presented 10 times in each of the sequences *hi* and *ih*. (ii) In Expts. I and II, daily morning and afternoon sessions 1 and 2, 3 and 4 etc. together comprised all 12 test pairs, while single sessions each comprised three *S* and three *L* or three *S* and three *SM*. All aliquots were brought to 70°F. for tasting in a laboratory conditioned to an air temperature of 72°F. and a relative humidity of 50%.

ANALYSIS OF RESULTS

Repeatability

Table I lists the frequency with which individuals recorded specified rankings in their 20 appraisals of each aliquot pair, $S2 > 1$ denoting ranking of *S*2 above *S*1, etc. Fig. 1 portrays the repetitive fluctuations in the 6-subject 6-comparison aggregate frequencies of these rankings recorded for each series of solutions. (In Expt. III each session, and in Expts. I and II the morning and afternoon sessions of each day together constituted a complete replicate). No evident progressive trend in these totals is visible. A numerical trend test was made by ordering the 20 replicate total frequencies of all 12, 12 and 6 specified rankings in each experiment, assigning the appropriate Fisher-Yates (7) normalized rank score to these ordered frequencies, and computing the variance associated and not associated with any linear trend in the resulting scores. (Under conditions of statistical stability these scores should of course have the characteristics of random normal deviates). Variance ratios of $F(1, 18) = 1.56$, 1.44 and 1.13 for Expts. I, II and III respectively and of $F(3, 54) = 1.45$ in the aggregate resulted. As the 5% points of $F(1, 18)$ and of $F(3,54)$ are 4.41 and 2.78, this analysis is likewise indicative of no progressive group trend.

As an index of the statistical stability of individuals' appraisals, Cochran's (5) *Q* for matched samples was computed from the replicate frequencies of specified rankings by each subject in each experiment.

The resulting Q are brought together in Table II. In the absence of heterogeneity, these individual Q should be distributed approximately as χ^2 with 19 d.f. (5% point 30.1), their sub-total for each experiment as χ^2 with 114 d.f. (5% point 140 approx.) and their total for all 3 experiments as χ^2 with 342 d.f. (5% point 386 approx.). In fact only

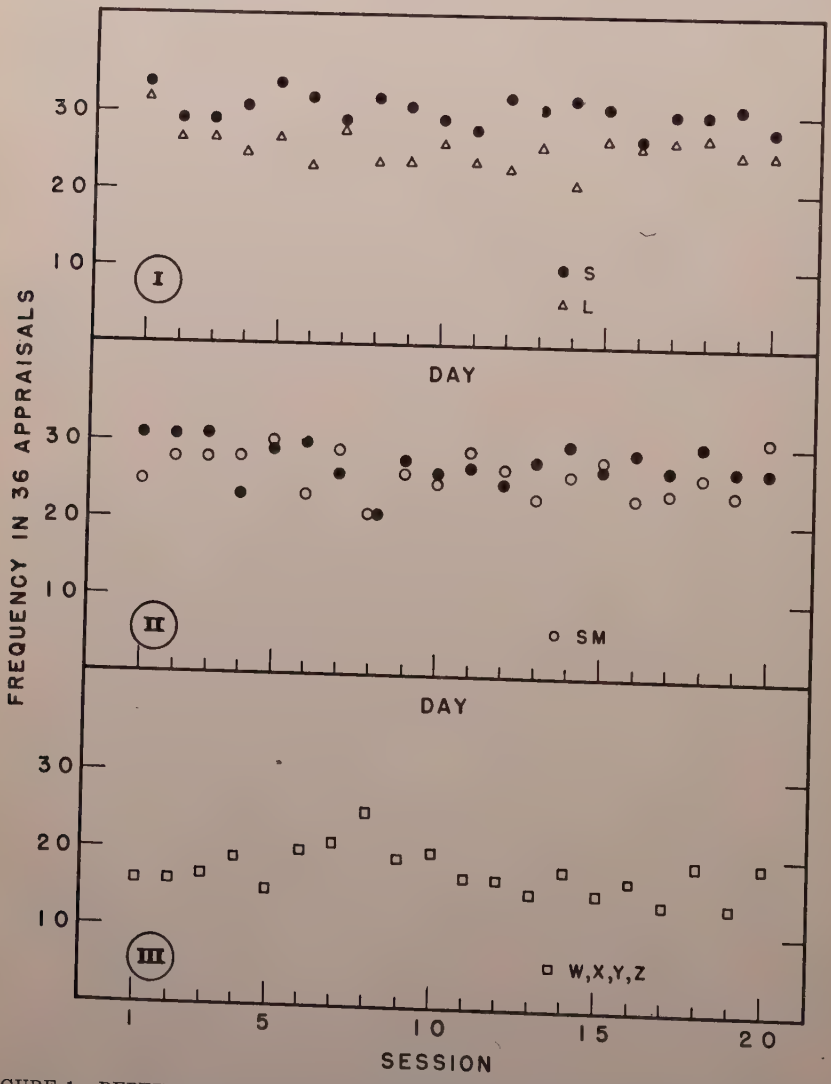


FIGURE 1. REPETITIVE FREQUENCY OF SPECIFIED RANKINGS IN EXPERIMENTS I, II AND III. *S*, SUCROSE ALONE; *L*, LACTOSE ALONE; *SM*, SUCROSE WITH OTHER PRIMARY TASTE STIMULI; *W*, *X*, *Y*, *Z*, QUALITATIVELY DIFFERENT TASTE MIXTURES.

TABLE II.

Indices Q of Discrepancies Between Replicates in Recorded Frequency of Specified Rankings by Individual Subjects

Experiment I		Experiment II		Experiment III	
Subject	Q (19 d.f.)	Subject	Q (19 d.f.)	Subject	Q (19 d.f.)
A	29.7	B	17.6	B	15.5
B	28.7	E	23.3	D	18.8
C	22.8	G	19.6	E	14.6
D	23.2	H	9.2	G	18.7
E	18.4	I	26.1	H	18.3
F	23.8	J	33.2	K	18.7
Total	146.6	Total	129.0	Total	104.6

one of the 18 individual Q listed, that for J in Expt. II, exceeds its 5% point. Subtotal Q for Expt. I attains approximately its 2% point, but as this is the largest of 3 subtotals it is suggestive rather than clearly indicative of correlated fluctuations in the rankings recorded in different replicates by the same subject. The other sub-total Q are entirely consistent with statistical independence of all rankings in Expts. II and III.

Bradley-Terry Rank Analysis

In Bradley and Terry's analysis (2), maximum likelihood estimates p_{ij} of the "ratings" π_{ij} from s replicate rankings of all pairings of m test samples are specified by the equations:

$$\frac{a_{ij}}{p_{ij}} - s \sum_{h \neq i} (p_{hj} + p_{ij})^{-1} = 0 \qquad (h, i = 1, 2, \cdots m)$$

$$\sum_{i} p_{ij} = 1$$

in which

$$a_{ij} = 2s(m - 1) - \sum_{h \neq i} \sum_{k=1}^{k=s} r_{ihjk} ,$$

r_{ihjk} being the rank assigned by the jth subject to the ith test sample in the kth replicate paired comparison of h and i. Approximate solution by iteration of these equations for the rank frequencies recorded in Table I gave the p_{ij} listed in Table III, and hence the Bradley-Terry expectations $f'_{ij} = 20p_{ij}/(p_{hj} + p_{ij})$ also listed in Table I.

TABLE III.

Maximum Likelihood Estimates of Bradley-Terry "Rating" Parameters for Individual Subjects

Test item	A	B	C	D	E	F
I.S. 1	.034	.007	.009	.024	.050	.005
2	.129	.031	.032	.053	.124	.024
3	.294	.058	.143	.221	.212	.208
4	.543	.904	.816	.702	.614	.763
I.L. 1	.150	.050	.040	.103	.065	.034
2	.184	.151	.148	.266	.183	.168
3	.168	.242	.291	.223	.343	.216
4	.498	.557	.521	.408	.409	.582
Test item	B	E	G	H	I	J
II.S. 1	.012	.025	.050	.113	.075	.091
2	.021	.093	.074	.120	.109	.138
3	.063	.138	.166	.305	.259	.285
4	.904	.744	.710	.462	.557	.486
II.SM. 1	.018	.107	.214	.033	.017	.106
2	.065	.142	.275	.100	.090	.167
3	.098	.260	.236	.300	.228	.206
4	.819	.491	.275	.567	.665	.521
Test item	B	D	E	G	H	K
III. W	.182	.454	.094	.176	.167	.180
X	.488	.236	.278	.196	.240	.171
Y	.075	.045	.509	.362	.435	.416
Z	.255	.265	.119	.266	.158	.233

Goodness of fit of the observed and expected frequencies was then tested as suggested by Bradley (4) by computing for each subject and test series the index of discrepancy $\sum(f - f')^2/f'$, this summation comprising in each instance the complementary frequencies of rankings $i > h$ and $h > i$. The results are shown in Table IV. On the Bradley-Terry assumptions, individual indices should be distributed approximately as χ^2 for 3 d.f. (5% point 7.82), their 6-subject subtotals approximately as χ^2 for 18 d.f. (5% point 28.9) and their total for all three experiments, here 84.2, approximately as χ^2 with 90 d.f. Clearly the

TABLE IV.

χ^2 (3 d.f.) from Individual Tests of Goodness of Fit of Recorded and Expected Rank Frequencies

Experiment I			Experiment II			Experiment III	
Subject	S	L	Subject	S	SM	Subject	WXYZ
A	2.99	1.11	B	2.72	5.17	B	2.24
B	5.04	0.80	E	1.14	3.40	D	2.53
C	2.21	3.23	G	5.34	2.05	E	0.63
D	4.17	1.67	H	1.31	2.57	G	6.07
E	3.82	1.74	I	1.99	5.78	H	1.36
F	6.83	0.46	J	3.49	0.75	K	1.55
Total (18 d.f.)	25.06	9.01	Total (18 d.f.)	15.99	19.72	Total (18 d.f.)	14.38

numerical values listed are individually and collectively indicative of excellent agreement between the observed and expected rank frequencies, notwithstanding the fact that several of the expectations f' complementary to those given in Table I are less than unity.

Difference Between Subjects

Homogeneity of individual reaction to the same paired comparison was tested directly by arraying the frequencies of ranking $i > h$ and $h > i$ recorded for each such comparison by the 6 participating subjects in the form of a 2×6 contingency table, computing χ^2 from these tables, and adding the results for each test series *I.S*, *I.L* etc. Table V shows these aggregate χ^2. Here again some of the expectations were small. In this instance however Haldane's (8) formulae permitted calculation of the exact moments of the conditional distribution of χ^2 for the marginal totals of each 2×6 table. The resulting aggregate $K_1(\chi^2)$ and $K_2(\chi^2)$ are accordingly also listed.

For the series *I.S*, *I.L* and *II.S*, χ^2 (90 d.f.) is 115.1, exceeding its expectation of 90.9 by 24.2, which is 1.88 times the corresponding standard deviation of $(166.88)^{1/2}$, namely 12.9. On the basis of upper tail probabilities this is suggestive, but only suggestive, of some real differences in individual discrimination of intensities of sweetness. There are evident differences between some of the corresponding total rank frequencies for *I.S* and *II.S* listed in the last column of Table I, but as the two tests were made at different times experimental factors (e.g. differences in tap water) may have contributed to these discrepancies.

Individual differences in respect of the *II.SM* comparisons were much more pronounced, and gave an aggregate χ^2 (30 d.f.) exceeding its expectation by 6.3 times the corresponding s.d. Indications from Tables I and III are that the discriminatory acuity for sweetness of subjects *E* and *G* was notably reduced by presence of the other three tastes; that of *B* and *J* was seemingly little affected, while that of *H* and *I* was if anything increased.

Individual differences in respect of comparisons III were still more pronounced, the aggregate χ^2 (30 d.f.) of 124.5 for these shown in Table V exceeding its expectation by 12.3 s.d. No two of the six

TABLE V.

χ^2 from Tests of Homogeneity of Frequency of Specified Rankings by Individual Subjects

Test items	χ^2 (30 d.f.)	K_1 (χ^2)	K_2 (χ^2)
I. S.	45.1	30.3	52.18
I. L.	35.1	30.3	57.58
II. S.	34.9	30.3	57.12
II. SM.	78.2	30.3	57.67
III. WXYZ.	124.5	30.3	58.44

participants rated all 4 test flavors in the same order. *Y* was rated first by 4 subjects but last by the other 2. *X* and *W* were both rated last by 2 subjects but first by another.

DISCUSSION

Repeatability and mutual independence of psychophysical reactions can seldom be taken for granted. In the present experiments, however, these conditions seem to have been reasonably approximated.

The Bradley-Terry model implies (3) that the probability of i being ranked above h in a single paired comparison by the jth subject may be specified by the logistic function

$$\frac{1}{4}\int_{-\ln(\pi_{ij}/\pi_{hj})}^{\infty} \operatorname{sech}^2 \frac{y}{2}\, dy$$

in which $y = \ln(\pi_{ij}/\pi_{hj})$ may be regarded as a measure of the degree of sensation evoked by the taste contrast (i, h). The generally good agreement of observed and hypothesized rank frequencies might then be regarded as consistent with existence of a specifically ordered continuum of degrees of sensation characteristic of each individual. For moderate frequencies the sigmoidal graph of a logistic is virtually in-

distinguishable from that of an integrated Gaussian function. Hence when this model is appropriate and stimulus differences are expressible in terms of a measured composition or process variable, knowledge of the appropriate transform $f(i, h) = y$ would permit repeated paired comparisons of a series of qualitatively similar test sample s to be regarded as organoleptic assays treatable statistically by the standard methods of logit or probit analysis (6).

In the foregoing experiments individual subjects reacted differentially to all but possibly the simplest flavor contrasts. Bradley and Terry (2) have noted that in these circumstances the "ratings" π_{ij} are functions of the subject as well as of the test material. Technological applications may then require numerical prediction of population reactions from those of small groups of laboratory subjects.

Opinion regarding commensurability of "degrees of preference" on a subjective continuum is currently divided (1, 10). It is therefore perhaps noteworthy that in these trials individuals' preference rankings of the qualitatively different test solutions were as consistent with the Bradley-Terry conceptual model as were the rankings of single flavor intensities.

ACKNOWLEDGEMENTS

The thanks of the author are due to the participating subjects, for sustained conscientious collaboration; also to N. T. Gridgeman and Elinor M. Zuckerman for technical supervision.

REFERENCES

1. Bartlett, F. C. Subjective judgments. *Nature, 166(4232)*: 984–985. 1950.
2. Bradley, R. A. and Terry, M. E. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika, 39*: 324–345. 1952.
3. Bradley, R. A. Some statistical methods in taste testing and quality evaluation. *Biometrics, 9*: 22–38. 1953.
4. Bradley, R. A. Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. *Biometrics 10*: 375–390, 1954.
5. Cochran, W. G. The comparison of percentages in matched samples. *Biometrika, 37*: 256–266. 1950.
6. Finney, D. J. *Statistical Methods in Biological Assay.* C. Griffin and Co. Ltd., London. 1952.
7. Fisher, R. A. and Yates, F. *Statistical Tables for Biological, Agricultural and Medical Research.* 3rd ed. Oliver and Boyd, London. 1948.
8. Haldane, J. B. S. The use of χ^2 as a test of homogeneity in a $(n \times 2)$-fold table when expectations are small. *Biometrika, 33*: 234–238. 1945.
9. Terry, M. E., Bradley, R. A. and Davis, L. L. New designs and techniques for organoleptic testing. *Food Technology, 6*: 250–254. 1952.
10. Thurstone, L. L. The prediction of choice. *Psychometrika, 10*: 237–253. 1945.

A NOTE ON MISSING-PLOT VALUES

J. A. Nelder
National Vegetable Research Station, Wellesbourne, Warwick, England

This note owes its origin to the discussion in *Biometrics* between Professor Snedecor and Dr. Williams on Query 96 (*Biometrics* Vol. 8 No. 4). In his original answer Professor Snedecor repeats what has often been written when he says 'The value of X is not intended as an estimate of the missing datum'. Nevertheless whether or not X is *intended* to be an estimate of the missing datum, it *is* an estimate of the missing datum, and an unbiased one where the mathematical model used is true. Thus, in the randomised block design, a commonly used model is that in which $x_{ij} = m + \beta_i + \tau_j + \epsilon_{ij}$ where x_{ij} is the observation in the jth treatment in the ith block, m is the mean, β and τ are (fixed) block and treatment constants respectively (subject to the usual constraints $\sum \beta = 0$, $\sum \tau = 0$) and ϵ_{ij} are the random residuals, mutually uncorrelated. If it is further assumed that the variance of ϵ_{ij} is constant, i.e. independent of i and j, then the formula for the missing plot value (y) in a randomised block experiment is $y = [rB + tT - G]/[(r - 1)(t - 1)]$, where r = no. of replicates, t = no. of treatments, B is the total of the remaining values in the block with the missing plot, T the similar total for the treatment, and G is the total of all existing values.

This value is the one obtained in a full least-squares analysis of the complete experiment, and its expectation is $m + \beta_i + \tau_j$, i.e. the mean value of the missing datum. Its variance is a minimum among linear unbiased estimates by the familiar properties of least squares. The variance of y for the randomised block design is given by

$$\text{var } (y) = \frac{(rt - 1)}{(r - 1)(t - 1)} \sigma^2$$

and for the $r \times r$ Latin square by

$$\text{var } (y) = \frac{3r - 2}{(r - 1)(r - 2)} \sigma^2$$

It is important to note that if the conditions postulated above do not hold, for example, if the variance of the ϵ_{ij} is a function of i and j, *then the missing plot value given above is no longer correct*, and does not lead to the correct analysis.

An 'impossible' value of y (such as a negative yield) or an unreasonably discrepant one can arise through sampling error, or it can arise through a failure of the mathematical model, and it may be of some importance to distinguish which is the cause in any particular instance. On the assumption that the missing plot has occurred at random in the experiment, we could use the estimated variance of the missing plot value to form a confidence interval in the usual way. If such a confidence interval for an essentially positive quantity contains no positive values, this is evidence against the negative value's being due to sampling error, and so in favour of its being due to a failure of the model. The practical value of this test is limited, however, by our knowing little about the expected value of the missing-plot value, other than that it must be positive if it is a yield or a count, etc. In Query 96, the test gives -6.64 ± 8.10 as the 95 per cent. confidence interval for y. Hence positive real values for y are not ruled out, and yet it is fairly certain from other evidence, as Dr. Williams points out, that a multiplicative model is much nearer the truth here than an additive one. Even when the test fails to discredit the sampling error hypothesis it is still hardly sufficient to suppose that the estimation of the error variance and the tests of significance will be necessarily valid. In Query 96, the variance of the catches in the treatment with the highest mean (B) is over 8 times the variance of the catches in the treatment with the lowest mean (A). The use of a common error variance for all comparisons will give considerably biased standard errors for some of the comparisons. Here, it happens not to matter, but this is nothing to take comfort from.

The analysis of variance receives rough treatment from some of its practitioners, and it is one of its merits that it can stand much of this without serious breakdown. However, it is as well not to try it too hard. An 'impossible' missing plot value is a danger signal, and should be taken heed of.

A NOTE ON TRUNCATED POISSON DISTRIBUTIONS

P. G. Moore
University College, London and Princeton University

1. In an observed Poisson distribution we have counts of the number of times that a certain event has occurred once, twice and so on in a set of N readings. From theoretical considerations we obtain the probability of just r occurrences of the event to be

$$p_r = \lambda^r e^{-\lambda}/r! \qquad r = 0, 1, 2, \cdots$$

where λ is the mean of the distribution and thus represents the average number of events to be expected per reading. An observed set of data, although Poisson in form, may exhibit various types of truncation due to the whole of the data not being available. If n_r be the observed frequency of just r occurrences then the main types of truncation are

(1) n_r for $r \geq s$ not observed.
(2) n_r for $r \leq k$ not observed.
(3) combination of (1) and (2).
(4) n_r for $r = v$ to $r = w$ inclusive not observed.

Truncations of Type 1 arise when there is difficulty in counting high numbers owing to inability to distinguish each individual or when the counting apparatus gives trouble for high counts. Type 2 occurs when it is desired to estimate the mean frequency of some event and only cases of more than k occurrences are reported. Type 3 occurs sometimes in botanical work where only quadrats with at least one individual in them are retained and high densities are difficult to count. Type 4 could occur when the observations have been muddled in some way.

Type 1 has been considered by Tippett [11] who derived the maximum likelihood solution, by Bliss [1] who gave tables to facilitate the solution of the equation and by Moore [6] who suggested a simple form of estimate needing no tables but having a larger standard error. Type 2, in the case where k is equal to zero, has been studied by David and Johnson [2] who gave the maximum likelihood solution, by Plackett [2] who gives a solution similar to that of Moore for Type 1 and Rider [9] who uses the first two incomplete moments. Cohen [3] has provided maximum likelihood estimates similar to those of Tippett for the cases of Types 2 — 4 and indicated how to solve the resulting equations by an

iterative or interpolatory procedure using a table of the Poisson distribution. In this note we will show that the type of estimate given in our earlier paper may be applied to any of Types 1 to 4 truncations by merely making a suitable change of the limits of summation. This quick form of estimate might be useful in any of the following situations

(a) when no extensive table of the Poisson distribution is available,
(b) when it is desired to obtain an estimate quickly without much computation,
(c) as a first approximation to use in the full maximum likelihood solutions of Tippett and Cohen.

It is, of course, only to be expected that the standard error of such an estimate will be larger than for the corresponding maximum likelihood solution.

2. For a Poisson distribution, truncated as in Type 1, we may write the identity

$$\sum_{r=0}^{s-1} r \frac{\lambda^r e^{-\lambda}}{r!} = \lambda \sum_{r=0}^{s-2} \frac{\lambda^r e^{-\lambda}}{r!}$$

and this suggests using as an estimate of λ the expression

Type 1 $$\hat{\lambda}_1 = \sum_{r=0}^{s-1} r n_r \Big/ \sum_{r=0}^{s-2} n_r$$

A similar form of argument for the other three types of truncation will give us as our respective estimators

Type 2 $$\hat{\lambda}_2 = \sum_{r=k+2}^{m} r n_r \Big/ \sum_{r=k+1}^{\infty} n_r$$

Type 3 $$\hat{\lambda}_3 = \sum_{r=k+2}^{s-1} r n_r \Big/ \sum_{r=k+1}^{s-2} n_r$$

Type 4 $$\hat{\lambda}_4 = \left\{\sum_{r=0}^{s-1} r n_r + \sum_{r=w+2}^{\infty} r n_r\right\} \Big/ \left\{\sum_{r=0}^{s-2} n_r + \sum_{r=w+1}^{\infty} n_r\right\}$$

Thus the only quantities used in the various estimators are r and rn_r.

3. As an example we will consider the following data due to Scrase [10].

r	0	1	2	3	4	5	6	7	8	Total
n_r	23	56	88	95	73	40	17	5	3	400

The problem is connected with the number of dust nuclei in the air and the data give the frequency distribution of the number of drops in a small volume of air that fall on to a stage in a chamber containing moisture and filtered air. The estimates obtained using the various formulae are tabulated in Table 1 where the estimate is given after the

TABLE 1.

Type 1	$s = 5$	3.088	$s = 6$	3.012	$s = 7$	2.963
Type 2	$k = 0$	2.955	$k = 1$	2.922	$k = 2$	2.803
Type 3	$k = 0$ $s = 7$	2.997	$k = 0$ $s = 6$	3.054	$k = 1$ $s = 7$	2.970
Type 4	$v = 4$ $w = 5$	3.000				

differing truncation details. For the full distribution the mean is 2.925 but Scrase is of the opinion that this mean is slightly high in that a number of zero counts were wrongly rejected as being due to the apparatus not working.

4. There are two ways of looking at the standard errors of the estimates. To show this let $\tilde{N}$ be the actual number of observations counted, that is N less those individuals that have been truncated. Then in obtaining the standard error we can regard either N or alternatively $\tilde{N}$ as being constant from sample to sample. The first situation was used by Tippett and Moore whilst the second, which appears to lead to simpler results, was used by David and Plackett. It is clear that to lay down hard and fast rules as to which is the correct one to use is impossible and situations of each type arise. For instance in particle counting where cells with high frequencies are not counted or in accident returns for some fixed population submitted to an authority, the situation with N constant seems plausible. On the other hand if the population is varying and we have just the data concerning actual accidents we would regard $\tilde{N}$ rather than N as fixed.

Assuming N to be fixed we can obtain asymptotic expressions for the standard errors of truncations of Types 1 and 2 fairly easily by the use of expectation techniques. They simplify to

$$\text{Var}\,(\hat{\lambda}_1) \sim \left\{\sum_{r=0}^{s-1} r^2 p_r - \left(\sum_{r=0}^{s-1} r p_r\right)^2\right\} \Big/ N\left(\sum_{r=0}^{s-2} p_r\right)^2$$

$$\text{Var}\,(\hat{\lambda}_2) \sim \left\{\sum_{r=k+2}^{\infty} r^2 p_r - \left(\sum_{r=k+2}^{\infty} r p_r\right)^2\right\} \Big/ N\left(\sum_{r=k+1}^{\infty} p_r\right)^2$$

The variance of $\hat{\lambda}_1$ has been discussed by Moore, but for $\hat{\lambda}_2$ we append a few values in Table 2 when k is equal to zero, one and two. To compare the case of k equal to zero with the figures given by Plackett we should multiply by $(1 - n_0)/N$ since his total is for the non-truncated portion only. On doing this we find our results comparable with his. In situations where the whole distribution is available and λ is estimated from the mean, N Var (mean) is simply λ. This enables us to see from Table 2 the loss of accuracy due to the truncation of the data.

TABLE 2.

λ	N Var $(\hat{\lambda}_2)$		
	$k = 0$	$k = 1$	$k = 2$
1.0	3.0846	11.8376	52.4339
1.5	3.4089	10.0824	32.8254
2.0	3.6256	9.1150	24.8133
2.5	3.8914	8.5522	20.6345
3.0	4.1250	8.0800	17.8645
3.5	4.3837	7.7107	15.8528
4.0	4.6772	7.4367	14.2889
5.0	5.3713	7.1754	12.0162
6.0	6.1941	7.2961	10.6001

5. The foregoing procedure may, with some small modification, be used for the case of truncation of the binomial population

$$(q + p)^n$$

where n is known but p is unknown and the number of zeros is missing. Cases of this kind arise, for example, in genetical work where it is desired to estimate the proportion of sibs who inherit some abnormality if the parents possess a certain gene. The presence of the gene is, however, only known if at least one of the sibs possesses the character in question. Hence the number of zeros in sibships of a given size n cannot be known and we must use the remaining information by itself to estimate p. Fisher [5] gave the maximum likelihood solution and

Finney [4] gave tables to reduce the solution to a fairly rapid iterative procedure. Following our previous method we are led to using

$$\sum_{r=2}^{n} \frac{rn_r}{n-r+1} \Big/ \sum_{r=1}^{n-1} n_r$$

as an estimator of p/q and from this equation we obtain

$$\sum_{r=2}^{n} \frac{rn_r}{n-r+1} \Big/ \left\{ \sum_{r=1}^{n-1} n_r + \sum_{r=2}^{n} \frac{rn_r}{n-r+1} \right\}$$

as an estimator for p.

To illustrate this procedure we consider the following data, due to Pearson [7], which give the number of sibs with human albinism in families of size five. Substituting in the formula given above we get 0.326 as our estimate of the unknown probability.

r	1	2	3	4	5
n_r	25	23	10	1	1

REFERENCES

[1] Bliss, C. I.: Estimation of the mean and its error from incomplete Poisson distributions. *Connecticut Agricultural Experiment Station Bulletin, 513*, 1948.

[2] David, F. N. and Johnson, N. L.: The truncated Poisson. *Biometrics, 8*, 275–285, 1952.

[3] Cohen, A. C., Jr.: Estimation of the Poisson parameter in truncated samples. *J. Amer. Stat. Assoc., 49*, 158–168, 1954.

[4] Finney, D. J.: The truncated binomial distribution. *Ann. Eugen. Lond. 14*, 319–328, 1949.

[5] Fisher, R. A.: The effects of method of ascertainment upon estimation of frequencies. *Ann. Eugen. Lond., 6*, 13–25, 1936.

[6] Moore, P. G.: The estimation of the Poisson parameter from a truncated distribution. *Biometrika, 39*, 247–251, 1952.

[7] Pearson, K.: A monograph on albinism in man. *Drapers Company Research Memoirs*, 1913.

[8] Plackett, R. L.: The truncated Poisson distribution. *Biometrics, 9*, 485–488, 1953.

[9] Rider, P. R.: Truncated Poisson distributions. *J. Amer. Stat. Assoc. 48*, 826–830, 1953.

[10] Scrase, F. J.: The sampling errors of the Aitken Nucleus Counter. *Quart. J. R. Met. Soc., 61*, 367–378, 1935.

[11] Tippett, L. H. C.: A modified method of counting particles. *Proc. Roy. Soc. Series A, 137*, 434–446, 1932.

QUERIES

GEORGE W. SNEDECOR, *Editor*

110 QUERY: Data on the calcium, phosphorous, and magnesium content of turnip leaves were collected in the following manner as reported in Southern Cooperative Series Bulletin No. 10, 1951, "Studies of Sampling Techniques and Chemical Analyses of Vegetables". "Duplicate [microchemical] analyses were made on each of four randomly-selected leaves from each of four turnip plants picked at random ··· . Duplicate determinations were made on each ash solution from a particular leaf ··· . The analyses of the two sets of ash solutions were made at different times."

To simplify the form of the questions to be asked, the analyses of variance for calcium and phosphorous are given.

ANALAYSES OF VARIANCE

Variance Source	Calcium		Phosphorous	
	Degrees of Freedom	Mean Square	Degrees of Freedom	Mean Square
Total	60		61	
Plants	3	6.202154**	3	0.056375**
Ashings	1	0.02945N.S.	1	0.000467N.S.
Plants × Ashings	3	0.033569**	3	0.000664N.S.
Leaves in Plants	12	0.605917**	12	0.035786**
Leaves × Ashings	11	0.013968**	11	0.000935N.S.
Duplicates	30	0.003560	31	0.000457

**Significant at the 1 per cent level of significance.
N.S. Not significant at either the 1 per cent level or the 5 per cent level of significance.

It is evident that the mean square for duplicate determinations must have been used as an estimate of error for testing the significance of Plants × Ashings, otherwise this term would not have been significant in the calcium analysis. Would not a more appropriate test have been made using Leaves [in plants] × Ashings as an estimate of error for Plants × Ashings?

In the case of testing for significance of Plants it is not quite so evident what source of variation was used as an estimate of error, but from the Phosphorous analysis it may be seen that of two possible

"contending" terms for error, Leaves in Plants and Plants × Ashings, each of which seems to possess the characteristics of a proper error term, Leaves in Plants was not used, otherwise Plants would not have been significant. How was it decided that Leaves in Plants would not be used as error for Plants? Or to make one generalized question out of the two, how does one select an appropriate error term in such an analysis as this?

ANSWER: This is a troublesome problem which arises when a factor under investigation is tried over two or more other factors which may be regarded as random. A number of articles shedding light on various phases of this problem have already appeared in this Journal: Wilm, H. G., *V*:1; Satterthwaite, F. E., *V*:2; Crump, S. L., *V*:2; Eisenhart, C., *V*:3; Cochran, W. G., *V*:7; and Ostle, Bernard, *V*:8 (Query).

The choice of a proper error term in any F test of significance depends upon the components of variance entering the two mean squares. The composition of any mean square can be determined only after the assumptions are made as to whether the factors under study are random or fixed.

In this particular case it would seem only reasonable, since the investigators were concerned primarily with sampling technique, and since they went to the trouble of specifying that both plants and leaves on plants were randomly selected, to assume (helped by the phraseology "Duplicate analyses $\cdots$") that the two ashings are also to be regarded as a random sample of a large number of ashings that might result from repeating the experiment many times on the same leaves by taking at successively later dates a new random sample from each leaf to be ashed and then analyzed in duplicate.

Under the assumption that all of the variables are random we may write the expectations of the mean squares for the various sources of variation as in Table 1.

For a case quite similar to this, but in which two of the factors were regarded as fixed, see Query 95, this Journal, Sept., 1952.

The proper mean square to use as an estimate of error for testing the significance of any source of variation is that mean square whose expectation is the same as the expectation of the mean square being tested except for that one component which is directly due to the source of variation being tested. This is really a test of the null hypothesis that the component due directly to the source being tested equals zero, for if we set this component equal to zero then both mean squares have the *same* expectation.

TABLE 1.

Expected Mean Squares of the Several Sources of Variation

Source of Variation	E(MS)
Plants—(P)	$\sigma_D^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2 + 4\sigma_{LP}^2 + 16\sigma_P^2$
Ashings—(A)	$\sigma_D^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2 + 32\sigma_A^2$
Plants × Ashings—(PA)	$\sigma_D^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2$
Leaves in Plants (LP)	$\sigma_D^2 + 2\sigma_{LPA}^2 + 4\sigma_{LP}^2$
Leaves in Plants × Ashings (LPA)	$\sigma_D^2 + 2\sigma_{LPA}^2$
Duplicates (D)	σ_D^2

To find, then, an appropriate error term for testing any particular mean square one looks for a mean square having the same expectation as has the mean square to be tested when the component due directly to the source being tested is set equal to zero.

For example, the expectation of Leaves in Plants × Ashings is $\sigma_D^2 + 2\sigma_{LPA}^2$. Setting σ_{LPA}^2, (the component due directly to Leaves in Plants × Ashings) equal to zero, there remains σ_D^2. The mean square having expectation σ_D^2 is Duplicates. Thus the mean square for duplicates is the appropriate error term for testing differences among Leaves in Plants × Ashings:

$$F = \frac{MS_{LPA}}{MS_D}$$

As you suggest, the appropriate error term for testing Plants × Ashings mean square is Leaves in Plants × Ashings mean square. The expectation of Plants × Ashings is $\sigma^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2$, and if the component due directly to Plants × Ashings (σ_{PA}^2) be set equal to zero, there remains $\sigma_D^2 + 2\sigma_{LPA}^2$ which is the expectation of Leaves in Plants × Ashings. For calcium:

$$F = \frac{MS_{PA}}{MS_{LPA}} = \frac{0.033569}{0.013968} = 2.40$$

On comparison with the tabled values $F = 2.40$ (with 3 and 11 degrees of freedom) is found to have a probability of occurrence greater than 0.05 due to chance alone, even when no additional component

exists in the numerator. Plants × Ashings is therefore *not* to be regarded as significant, contrary to the analysis quoted.

The test of significance for Plants is more complex. The expectation of the mean square for plants is $\sigma_D^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2 + 4\sigma_{LP}^2 + 16\sigma_P^2$. Setting the component directly due to Plants (σ_P^2) equal to zero the expectation is $\sigma_D^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2 + 4\sigma_{LP}^2$. Examination of the expectations of the several mean squares of the analysis reveals that none of them has this expectation. However, a mean square with this expectation may be constructed as follows:

$$MS'_E = MS_{PA} + MS_{LP} - MS_{LPA}$$

If we set down the expectations of the various mean squares of the constructed mean square and perform the specified operations we have:

$$MS'_E = (\sigma_D^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2) + (\sigma_D^2 + 2\sigma_{LPA}^2 + 4\sigma_{LP}^2) - (\sigma_D^2 + 2\sigma_{LPA}^2)$$
$$= \sigma_D^2 + 2\sigma_{LPA}^2 + 8\sigma_{PA}^2 + 4\sigma_{LP}^2$$

Thus the constructed mean square has exactly the expectation necessary to serve as error for testing significances of differences among Plants. It is used in the same manner as any other mean square:

$$F' = \frac{MS_P}{MS'_E}$$

The degrees of freedom for the constructed mean square are not known but may be approximated by a method due to Satterthwaite:

$$\text{d.f.}'_E = \frac{(MS'_E)^2}{\dfrac{(MS_{PA})^2}{\text{d.f.}_{PA}} + \dfrac{(MS_{LP})^2}{\text{d.f.}_{LP}} + \dfrac{(MS_{LPA})^2}{\text{d.f.}_{LPA}}}$$

It has been pointed out however that the presence of the minus sign in the construction of MS'_E, especially if the degrees of freedom for the various mean squares are small so that the estimates might not be the most reliable, could possibly lead to a negative estimate of MS'_E in which case the approximating F distribution would not be too good. To overcome this drawback it has been proposed (Cochran) that the test be preformed by the calculation of:

$$F'' = \frac{MS''_P}{MS''_E} = \frac{MS_P + MS_{LPA}}{MS_{PA} + MS_{LP}},$$

which gives two mean squares whose expectations are the same under the null hypothesis that $\sigma_P^2 = 0$. Degrees of freedom must now be

approximated for both numerator and denominator of the F ratio.

$$\text{d.f.}''_P = \frac{(MS''_P)^2}{\frac{(MS_P)^2}{\text{d.f.}_P} + \frac{(MS_{LPA})^2}{\text{d.f.}_{LPA}}}$$

$$\text{d.f.}''_E = \frac{(MS''_E)^2}{\frac{(MS_{PA})^2}{\text{d.f.}_{PA}} + \frac{(MS_{LP})^2}{\text{d.f.}_{LP}}}$$

Solving for both F' and F'' using the analysis of phosphorous content we find:

(1) $$F' = \frac{0.056375}{0.000664 + 0.035786 - 0.000935} = \frac{0.056375}{0.035515} = 1.59$$

$$\text{d.f.}'_E = \frac{(0.035515)^2}{\frac{(0.000664)^2}{3} + \frac{(0.035786)^2}{12} + \frac{(0.000935)^2}{11}}$$

$$= \frac{0.001,261,315}{0.000,106,946} = 12$$

Assuming now that F' with 3 and 12 degrees of freedom is distributed as F we look in the Table of F and see that values as large as 1.59 (3, 12 d.f.) occur in more than 10% of trials due to chance alone, so that plants are not to be regarded as differing significantly in phosphorus content.

(2) $$F'' = \frac{0.056375 + 0.000935}{0.000664 + 0.035786} = \frac{0.057310}{0.036450} = 1.57$$

$$\text{d.f.}''_P = \frac{(0.057310)^2}{\frac{(0.056375)^2}{3} + \frac{(0.000935)^2}{11}} = \frac{0.003,284,436}{0.001,059,459} = 3.1 = 3$$

$$\text{d.f.}''_E = \frac{(0.036450)^2}{\frac{(0.000664)^2}{3} + \frac{(0.035786)^2}{12}} = \frac{0.0001,328,602}{0.000,106,867} = 12.4 = 12$$

We now have $F'' = 1.57$ (3, 12 d.f.) from which we reach the same conclusion as before, these data are insufficient to say that differences in phosphorous content exist among plants, contrary again to the quoted analysis.

E. F. Schultz, Jr.

QUERY: Comment on Queries Nos. 96 and 103.

111 I find myself in partial agreement and partial disagreement with both querist and answerer in Query 103. The issues raised by this query seem important enough to warrant further discussion. The main point at issue is the usefulness of the logarithmic number of insects. This is by no means belittled by the fact that both counted and logarithmic values yield significance. The question is not the existence of insight into the data but the extent of the insight.

We may inquire into this by examining the values in more detail. When we do this, the logarithmic values produce curious suggestions concerning the 12 trap locations—suggestions which were covered up in the counted values. If these suggestions correspond to real features in the placement of the 12 platforms, there is no doubt that the logarithmic values have produced more insight. (And even if the suggested features are not known to be real, it is highly likely that they represent extra insight.)

These suggestions may be obtained by comparing individual traps with the means of other traps on the same platform. (They were actually observed by calculating apparent interactions of the form

(individual value)—(row mean)—(column mean) + (grand mean)

but we shall forego these in the interests of space.)

Reps	*A compared with mean of B, C, and D*	
	counts	*logarithms*
1, 2, 3, 4	−47, −32, −32, −13	−.82, −.85, −.68, −.71
5, 7, 9, 11	−35, −19, (−22), −10	−.45, −.43, (−.45), −.37
6, 8, 10, 12	−10, −13, −5, −21	−.18, −.18, −.09, −.26

The clear distinction between the three sets of 4 replications shown in the differences of logarithms is dimly, if at all, perceptible in the differences of counts. Here the logarithm has earned its salt—not just because of equal or unequal weighting—but through clarification of the relations of individual values.

In passing, let me record my support of *both* the statement by Snedecor that the supplied value 'is merely a number to be put in the empty space' for convenience *and* the implication by Williams that the appearance of unusual supplied values (negative, very large, or what

have you) should always be taken as a warning and a basis for careful consideration of how the analysis used relates to reality.

Incidentally, application of the computations of "One degree of freedom for nonadditivity" (Biometrics 5 (1949) 232–242)) to both the counts and the logarithms yields the following results:

		MS for Counts	MS for Logarithms
Nonadditivity	1	668	.014
Balance	32*	149	.036

(*1 supplied value)

The test for nonadditivity of counts is significant at 5%, while the test for nonadditivity of logarithms provides no slightest ground for suspicion. (The ratios of row MS and column MS to balance MS were each increased by a factor of about 1.5 by changing from counts to logarithms.) Thus an available test could have been applied to this rather restricted body of data to indicate the need of a transformation.

JOHN TUKEY

Editorial Comment:

On receipt of this penetrating observation, I asked querist for any information he might have that would serve to explain the peculiar groupings in the data. None was forthcoming, but three more batches of data from the same traps were sent. These were examined in the manner suggested by Dr. Tukey. In all four batches, suspension platforms 1 and 4 showed above average differences between the catch in trap A and the mean catch in traps B, C and D. On suspension 8, all differences were notably small.

Querist sent a map of the locations of the traps but these seemed to have no relation to the peculiarities observed in the data.

I am inclined to think that Dr. Tukey, by shrewd examination of the data, has discovered faults of construction in the traps.

THE BIOMETRIC SOCIETY

Finances for 1953. On recommendation of the Finance Committee, the financial reports of the Society will be printed in the future in *BIOMETRICS* instead of being sent separately to each member as in the past. The following are the audited statements for 1953, somewhat condensed.

BIOMETRICS

Income

From 1952, reserve from sale of No. 1 in Vols. 3 and 7	$ 1100.00	
Balance from 1952 books	3758.21	$ 4858.21
Member subscriptions		
ASA, 364	1447.00	
Biometric Society, 1117	3992.75	5439.75
Non-member subscriptions, 559		3914.75
Sale of back issues		
Vols. 1–5	$ 1261.75	
Vols. 6–9	1109.93	
Vol. 3, No. 1, 121; Vol. 7, No. 1, 20	221.50	2593.18
Sale of reprints		963.81
Total Income		$17769.70

Expenses

Overpayments, cancellations		$ 77.00
Bellagio Conference, offprints		85.46
Envelopes, stationery, supplies		432.90
Postage		358.18
Institute of Statistics, Editorial Management		1000.00
ASA, ½ net profit on sale of vols. 1–5		568.37
Wm. Byrd Press		
Production of Biometrics, Vol. 9, 1953	$ 7861.27	
Offprints of Dec. 1952 to Sept. 1953	1071.07	8932.34
Total Expenses		$11454.25
Balance		$ 6315.45

Office of the Secretary-Treasurer

Income

Subscriptions—1952		$ 286.50	
1953		3466.75	$ 3753.25
Dues—1952		$ 91.75	
1953		2155.75	$ 2247.50
Sustaining Members, 1953			800.00
Back dues and subscriptions			35.00
Reprints and Back issues	$ 67.60		
Striplist and adjustments	34.14		
Overpayments	39.50	141.24	
Less credits taken for overpayments made prior to 1953		48.25	92.99
Total Income			$ 6928.74

Expenses

	Directory	*General*	*Total*	
BIOMETRICS		$4354.69	$ 4354.69	
Salaries	$ 523.08	1191.56	1714.64	
Printing	716.11	88.84	804.95	
Stationery and supplies	33.86	67.42	101.28	
Postage	77.77	213.98	291.75	
Telephone		84.38	84.38	
Special services	5.63	75.75	81.38	
Depreciation		230.91	230.91	
Miscellaneous		26.53	26.53	
Total Expense	$1356.45	$6773.31		$ 8129.76
Excess of Expense over Income				$ 1201.02

Assets and Liabilities as of December 31, 1953

Assets

Cash on hand (including $35.00 petty cash)	$ 2938.91	
Office Equipment	230.89	
1953 dues and subscriptions in transit	452.63	
		$ 3622.43

Liabilities

Dues and Subscriptions for 1954	$ 1871.00	
Surplus of assets over liabilities	1751.43	
		$ 3622.43

Surplus of assets over liabilities, Dec. 31st, 1952	$ 2635.92
Surplus of assets over liabilities, Dec. 31st, 1953	$ 1751.43
Decrease, 1953	−$884.49

WNAR. The Western North American Region met on June 19 at the California Institute of Technology in Pasadena. A joint session with the Institute of Mathematical Statistics considered social and medical applications with papers by Lester Breslow on "Occupations and cigarette smoking as factors in lung cancer" and by A. W. Marshall on "A study of the onset of mental disease from admission data." The meeting then featured a special invited address by E. C. Hammond on "The problem of establishing cause and effect relationships in the etiology of chronic diseases." The afternoon program of contributed papers, by D. C. Joseph on "Morphological variations in a small marine fish (*Bairdiella icistius*) planted in the Salton Sea" and by John Radovich on "Estimating the size of the population of the Pacific sardine (*Sardinope caerulea*)", was followed by an informal symposium on methods of teaching biostatistics to medical students, nurses, sanitarians, zoologists, ichthyologists, and other ologists with F. H. Weymouth presiding and M. Elveback, C. E. Hopkins, L. E. Moses, W. J. Dixon and D. G. Chapman as participants.

The annual business meeting of the Region was also held on June 19 with the Regional President, D. G. Chapman, in the chair. After a discussion of meetings proposed for December and the next spring, the following were elected as Regional Officers for 1955: President—W. J. Dixon, Secretary-Treasurer—Elizabeth Vaughan, Regional Committee members—D. G. Chapman and B. M. Bennett.

A meeting of the WNAR of The Biometric Society will be held on the Berkeley Campus of the University of California, December 27–29, 1954, in conjunction with meetings of the American Association for the Advancement of Science, the first part of the Third Berkeley Symposium, the Institute of Mathematical Statistics, and the American Association for the Advancement of Science. Symposiums on the Design of Experiments in fisheries; Statistics in Biology and Genetics; and Statistics in Medicine and Public Health are being organized. Prof. Lincoln Moses of Stanford University, California, is chairman of the program committee.

Région pour la Belgique et le Congo Belge a eu le grand plaisir de recevoir le 7 juillet dans les locaux de la Fondation Universitaire, Bruxelles, M. Phillipe F. Bourdeaux, Ph.D., Ingénieur des Eaux et Forêts A.I.G.^{x}, Assistant-Professor au North Carolina State College, Raleigh, U.S.A. Le sujet du discours de Dr. Bourdeaux etait "Methodes biometriques dans l'analyse de la vegetation." La Conférence etait suivie d'une discussion active.